DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE

Publication number: WO03020118
Publication date: 2003-03-13
Inventor: MCCARTHY JEANETTE (US)
Applicant: VITIVITY INC (US); MCCARTHY JEANETTE (US)
Classification:
- International: ***C07H21/00; C07H21/04; C12Q1/68; G01N33/48; G01N33/50; G06F19/00;*** **A61B; C07H21/00; C12Q1/68; G01N33/48; G01N33/50; G06F19/00;** (IPC1-7): A61B
- European: C12Q1/68M6
Application number: WO2002US28113 20020904
Priority number(s): US20010317178P 20010905; US20010329958P 20011016; US20010017724 20011214

Also published as:
WO03020118 (A3)
US2003099958 (A1)

Cited documents:
XP002977124
XP002977125
XP002977126
XP002977127
XP002971567

Report a data error here

Abstract of **WO03020118**

The present invention is based at least in part on the discovery of polymorphisms within the thrombospondin 2 (THBS2) gene, the angiotensin converting enzyme 1 (ACE), and the beta fibrinogen (FGB) gene. Accordingly, the invention provides nucleic acid molecules having a nucleotide sequence of an allelic variant of a THBS2, ACE, or FGB gene. The invention also provides methods for identifying specific alleles of polymorphic regions of a THBS2, ACE, or FGB gene, methods for determining whether a subject is or is not at risk of developing a disease which is associated with a specific allele of a polymorphic region of a THBS2, ACE, or FGB gene, e.g., a vascular disease, based on detection of polymorphisms within the THBS2, ACE, or FGB gene, and kits for performing such methods. The invention further provides methods for classifying a subject who is or is not at risk for developing, a vascular disease or disorder as a candidate for a particular clinical course of therapy or a particular diagnostic evaluation.

---

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 March 2003 (13.03.2003)

PCT

(10) International Publication Number
WO 03/020118 A2

(51) International Patent Classification[7]: A61B

(21) International Application Number: PCT/US02/28113

(22) International Filing Date:
4 September 2002 (04.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/317,178	5 September 2001 (05.09.2001)	US
60/329,958	16 October 2001 (16.10.2001)	US
10/017,724	14 December 2001 (14.12.2001)	US

(71) Applicant *(for all designated States except US)*: **VITIVITY, INC.** [US/US]; 1 Kendall Square, Building 700, Cambridge, MA 02139 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **MCCARTHY, Jeanette** [US/US]; 3625 Dupont Street, San Diego, CA 92106 (US).

(74) Agents: **SMITH, DeAnn, F.**; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/020118 A2

(54) Title: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE

(57) Abstract: The present invention is based at least in part on the discovery of polymorphisms within the thrombospondin 2 (THBS2) gene, the angiotensin converting enzyme 1 (ACE), and the beta fibrinogen (FGB) gene. Accordingly, the invention provides nucleic acid molecules having a nucleotide sequence of an allelic variant of a THBS2, ACE, or FGB gene. The invention also provides methods for identifying specific alleles of polymorphic regions of a THBS2, ACE, or FGB gene, methods for determining whether a subject is or is not at risk of developing a disease which is associated with a specific allele of a polymorphic region of a THBS2, ACE, or FGB gene, *e.g.*, a vascular disease, based on detection of polymorphisms within the THBS2, ACE, or FGB gene, and kits for performing such methods. The invention further provides methods for classifying a subject who is or is not at risk for developing, a vascular disease or disorder as a candidate for a particular clinical course of therapy or a particular diagnostic evaluation.

Diagnosis and Treatment of Vascular Disease

Related Applications

This application claims priority to U.S. Patent Application No. 10/017,724, filed December 14, 2001 (pending), which claims priority to U.S. Provisional Application Serial No. 60/317,178, filed on September 5, 2001, and to U.S. Provisional Application Serial No. 60/329,958, filed on October 16, 2001, the contents of which are incorporated herein in their entirety by reference.

Background of the Invention

Cardiovascular disease is a major health risk throughout the industrialized world. Coronary artery disease (CAD), or atherosclerosis, involves the progressional narrowing of the arteries due to a build-up of atherosclerotic plaque. Myocardial infarction (MI), *e.g.*, heart attack, results when the heart is damaged due to reduced blood flow to the heart caused by the build-up of plaque in the coronary arteries.

Coronary artery disease, the most prevalent of cardiovascular diseases, is the principal cause of heart attack, stroke, and gangrene of the extremities, and thereby the principle cause of death in the United States. Coronary artery disease, or atherosclerosis, is a complex disease involving many cell types and molecular factors (described in, for example, Ross, 1993, *Nature* 362: 801-809). The process, in normal circumstances a protective response to insults to the endothelium and smooth muscle cells (SMCs) of the wall of the artery, consists of the formation of fibrofatty and fibrous lesions or plaques, preceded and accompanied by inflammation. The advanced lesions of atherosclerosis may occlude the artery concerned, and result from an excessive inflammatory-fibroproliferative response to numerous different forms of insult. Injury or dysfunction of the vascular endothelium is a common feature of may conditions that predispose a subject to accelerated development of atherosclerotic cardiovascular disease. For example, shear stresses are thought to be responsible for the frequent occurrence of atherosclerotic plaques in regions of the circulatory system where turbulent blood flow occurs, such as branch points and irregular structures.

The first observable event in the formation of an atherosclerotic plaque occurs when blood-borne monocytes adhere to the vascular endothelial layer and transmigrate through to the sub-endothelial space. Adjacent endothelial cells at the same time produce oxidized low density lipoprotein (LDL). These oxidized LDLs are then taken up in large amounts by the monocytes through scavenger receptors expressed on their surfaces. In contrast to the regulated pathway by which native LDL (nLDL) is taken up by nLDL specific receptors, the scavenger pathway of uptake is not regulated by the monocytes.

These lipid-filled monocytes are called foam cells, and are the major constituent of the fatty streak. Interactions between foam cells and the endothelial and SMCs which surround them lead to a state of chronic local inflammation which can eventually lead to smooth muscle cell proliferation and migration, and the formation of a fibrous plaque.

Such plaques occlude the blood vessel concerned and, thus, restrict the flow of blood, resulting in ischemia. Ischemia is a condition characterized by a lack of oxygen supply in tissues of organs due to inadequate perfusion. Such inadequate perfusion can have a number of natural causes, including atherosclerotic or restenotic lesions, anemia, or stroke. Many medical interventions, such as the interruption of the flow of blood during bypass surgery, for example, also lead to ischemia. In addition to sometimes being caused by diseased cardiovascular tissue, ischemia may sometimes affect cardiovascular tissue, such as in ischemic heart disease. Ischemia may occur in any organ, however, that is suffering a lack of oxygen supply.

One of the most important risk factors for coronary artery disease is a familial history. Although family history subsumes both genetic and shared environmental factors, studies suggest that CAD has a very strong genetic component (Marenberg, *et al.* (1994) *NEJM* 330:1041). Despite the importance of family history as a risk factor for CAD, it's incomplete genetic basis has not been elucidated. Therefore, the identification of genes which are involved in the development of CAD and MI would be beneficial.

It would thus be beneficial to identify polymorphic regions within genes which are associated with a vascular disease or disorder, such as coronary artery disease or myocardial infarction. It would further be desirable to provide prognostic, diagnostic, pharmacogenomic, and therapeutic methods utilizing the identified polymorphic regions.

Summary of the Invention

The present invention is based, at least in part, on the identification of polymorphic regions within the thrombospondin 2 (THBS2) gene, angiotensin converting enzyme 1 (ACE) gene, and the beta fibrinogen (FGB) gene which are associated with specific diseases or disorders, including vascular diseases or disorders. In particular, single nucleotide polymorphisms (SNPs) in these genes which are associated with premature coronary artery disease (CAD) (or coronary heart disease) and myocardial infarction (MI) have been identified. SNPs in these genes, as identified herein, singly or in combination, can be utilized to predict, in a subject, a decreased risk for developing a vascular disease, *e.g.*, CAD and/or MI.

The SNPs identified herein may further be used in the development of new treatments for vascular disease based upon comparison of the variant and normal versions of the gene or gene product (*e.g.*, the reference sequence), and development of cell-culture based and animal models for research and treatment of vascular disease. The invention further relates to novel compounds and pharmaceutical compositions for use in the diagnosis and treatment of such disorders. In preferred embodiments, the vascular disease is CAD or MI.

The polymorphisms of the invention may thus be used, both singly or in combination, in prognostic, diagnostic, and therapeutic methods. For example, the polymorphisms of the invention can be used to determine whether a subject is or is not at risk of developing a disease or disorder associated with a specific allelic variant of a THBS2, ACE, or FGB polymorphic region, *e.g.*, a disease or disorder associated with aberrant THBS2, ACE, or FGB activity, *e.g.*, a vascular disease or disorder such as CAD or MI.

The invention thus relates to isolated nucleic acid molecules and methods of using these molecules. The nucleic acid molecules of the invention include specific THBS2, ACE, or FGB allelic variants which differ from the reference THBS2, ACE, or FGB sequences set forth in SEQ ID NO:1 (GI 307505), SEQ ID NO:3 (GI 13027555), or SEQ ID NO:5 (GI 182597), respectively, or a portion thereof. The preferred nucleic acid molecules of the invention comprise THBS2, ACE, or FGB polymorphic regions or

portions thereof having the polymorphisms shown in Tables 1, 4, and 6 (corresponding to SEQ ID NOs.:7, 8, 9, 10, and 11), polymorphisms in linkage disequilibrium with the polymorphisms shown in Tables 1, 4, and 6, and combinations thereof. Nucleic acids of the invention can function as probes or primers, *e.g.*, in methods for determining the allelic identity of a THBS2, ACE, or FGB polymorphic region in a nucleic acid of interest.

The nucleic acids of the invention can also be used, singly or in combination, to determine whether a subject is or is not at risk of developing a disease associated with a specific allelic variant of a THBS2, ACE, or FGB polymorphic region, *e.g.*, a disease or disorder associated with aberrant THBS2, ACE, or FGB activity, *e.g.*, a vascular disease or disorder such as CAD or MI. The nucleic acids of the invention can further be used to prepare THBS2, ACE, or FGB polypeptides encoded by specific alleles, such as mutant (variant) alleles. Such polypeptides can be used in therapy. Polypeptides encoded by specific THBS2, ACE, or FGB alleles, such as variant THBS2, ACE, or FGB polypeptides, can also be used as immunogens and selection agents for preparing, isolating or identifying antibodies that specifically bind THBS2, ACE, or FGB proteins encoded by these alleles. Accordingly, such antibodies can be used to detect variant THBS2, ACE, or FGB proteins.

There are two preferred SNPs in the THBS2 gene. One polymorphism found in the THBS2 gene in the population screened is a change from a thymidine (T) to a guanine (G), or the complement thereof, in the THBS2 gene at residue 3949 of the reference sequence GI 307505 (polymorphism ID No. G5755e5). A second polymorphism in the THBS 2 gene is a change from a thymidine (T) to a cytidine (C), or the complement thereof, at residue 4476 of the reference sequence GI 307505 (polymorphism ID No. G5755e9). These polymorphisms are located in the 3' untranslated region (UTR) of the THBS2 gene, and therefore do not result in a change in the amino acid sequence of the THBS2 protein.

There is one preferred SNP in the ACE gene. This SNP, identified herein as G765u2, is a change from an adenine (A) to a guanine (G), or the complement thereof, at nucleotide residue 86408 of the ACE reference sequence GI 13027555. This SNP is a

"silent" variant. That is, it does not result in a change in the amino acid sequence of the ACE protein.

There are two preferred SNPs in the FGB gene. One SNP, referred to herein as FGBu1, is a change from a cytidine (C) to a thymidine (T), or the complement thereof, at nucleotide residue 5119 of the FGB reference sequence GI 182597. This SNP is a silent variant. The second SNP, FGBu4, is a change from a guanine (G) to an adenine (A), or the complement thereof, at nucleotide residue 8059 in the reference sequence GI 182597. This polymorphism is a missense variation which results in a change from an arginine (R) to a lysine (K) in the amino acid sequence of FGB (SEQ ID NO:6) at amino acid residue 478.

The nucleic acid molecules of the invention can be double- or single-stranded. Accordingly, in one embodiment of the invention, a complement of the nucleotide sequence is provided wherein the polymorphism has been identified. For example, where there has been a single nucleotide change from a thymidine to a cytidine in a single strand, the complement of that strand will contain a change from an adenine to a guanine at the corresponding nucleotide residue. The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a polymorphism of the present invention or to its complement.

The polymorphisms of the present invention, either singly, in combination with each other, or in combination with previously identified polymorphisms, are shown herein to be associated with specific disorders, *e.g.,* vascular diseases or disorders. Examples of vascular diseases or disorders include, without limitation, atherosclerosis, coronary artery disease (CAD), myocardial infarction (MI), ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

The invention further provides vectors comprising the nucleic acid molecules of the present invention; host cells transfected with said vectors whether prokaryotic or eukaryotic; and transgenic non-human animals which contain a heterologous form of a functional or non-functional THBS2, ACE, or FGB allele described herein. Such a transgenic animal can serve as an animal model for studying the effect of specific THBS2, ACE, or FGB allelic variations, including mutations, as well as for use in drug screening and/or recombinant protein production.

In another preferred embodiment, the method comprises determining the nucleotide content of at least a portion of a THBS2, ACE, or FGB gene, such as by sequence analysis. In yet another embodiment, determining the molecular structure of at least a portion of a THBS2, ACE, or FGB gene is carried out by single-stranded conformation polymorphism (SSCP). In yet another embodiment, the method is an oligonucleotide ligation assay (OLA). Other methods within the scope of the invention for determining the molecular structure of at least a portion of a THBS2, ACE, or FGB gene include hybridization of allele-specific oligonucleotides, sequence specific amplification, primer specific extension, and denaturing high performance liquid chromatography (DHPLC). In at least some of the methods of the invention, the probe or primer is allele specific. Preferred probes or primers are single stranded nucleic acids, which optionally are labeled.

The methods of the invention can be used for determining the identity of a nucleotide or amino acid residue within a polymorphic region of a human THBS2, ACE, or FGB gene present in a subject. For example, the methods of the invention can be useful for determining whether a subject is or is not at risk of developing a disease or condition associated with a specific allelic variant of a polymorphic region in the human THBS2, ACE, or FGB gene, *e.g.*, a vascular disease or disorder.

In one embodiment, the disease or condition is characterized by an aberrant THBS2, ACE, or FGB activity, such as aberrant THBS2, ACE, or FGB protein level, which can result from aberrant expression of a THBS2, ACE, or FGB gene. The disease or condition can be CAD, MI, or another vascular disease. Accordingly, the invention provides methods for predicting a subject's risk for developing a vascular disease associated with aberrant THBS2, ACE, or FGB activity. In a preferred embodiment, a subject having "pattern 1," which comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT), or the complement thereof, or "pattern 2", which comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG), or the complement thereof, is at a approximately 3-fold decreased odds of vascular disease compared to all other combinations of genotypes at these two loci.

In another preferred embodiment, a subject having one copy of an A and one copy of a G at nucleotide 86408 of the ACE reference sequence GI 13027555 (AG genotype), or the complement thereof, is at a decreased risk for vascular disease relative to persons with other genotypes for this SNP (*e.g.*, AA or GG genotypes).

In yet another preferred embodiment, a subject having two copies of a T at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is at a ~ 3-fold decreased risk for vascular disease relative to persons with the CC genotype. A subject having one copy of a T and one copy of a C at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is also at a decreased risk for vascular disease relative to persons with the CC genotype.

In still another preferred embodiment, a subject having two copies of an A at nucleotide residue 8059 of the FGB reference sequence GI 182597, or the complement thereof, is at a ~3-fold decreased risk for vascular disease relative to persons with the GG genotype. A subject having one copy of an A and one copy of a G at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is also at a decreased risk for vascular disease relative to persons with the GG genotype (see Example 1).

Additionally, the invention provides a method of identifying a subject who is or is not susceptible to a vascular disorder, which method comprises the steps of i) providing a nucleic acid sample from a subject; and ii) detecting in the nucleic acid sample the presence or absence of a THBS2, ACE, or FGB gene polymorphism, or both in combination, that correlate with the vascular disorder with a P value less than or equal to 0.05.

The invention further provides forensic methods based on detection of polymorphisms within the THBS2, ACE, or FGB gene.

The invention also provides probes and primers comprising oligonucleotides, which correspond to a region of nucleotide sequence which hybridizes to at least 6 consecutive nucleotides of the sequence set forth as SEQ ID NOs.:7, 8, 9, 10, and 11 or to the complement of the sequences set forth as SEQ ID NOs.:7, 8, 9, 10, and 11, or naturally occurring mutants or variants thereof. In preferred embodiments, the

probe/primer further includes a label attached thereto, which is capable of being detected.

A kit of the invention can be used, *e.g.*, for determining whether a subject is or is not at risk of developing a disease associated with a specific allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene, *e.g.*, a vascular disease, *e.g.*, CAD or MI. In a preferred embodiment, the invention provides a kit for determining whether a subject is or is not at risk of developing a vascular disease such as, for example, atherosclerosis, CAD, MI, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. The kit of the invention can also be used in selecting the appropriate clinical course of clinical treatment to a subject to treat a disease or condition, such as a disease or condition set forth above. Thus, determining the allelic variants of THBS2, ACE, or FGB polymorphic regions of a subject can be useful in predicting how a subject will respond to a specific drug, *e.g.*, a drug for treating a disease or disorder associated with aberrant THBS2, ACE, or FGB, *e.g.*, a vascular disease or disorder.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Figures

Figure 1 depicts the nucleotide sequence corresponding to reference sequence GI 307505 (SEQ ID NO:1) for the THBS2 gene.

Figure 2 depicts the amino acid sequence corresponding to reference GI 4507487 (SEQ ID NO:2) for the THBS2 protein.

Figure 3 depicts the nucleotide sequence corresponding to reference sequence GI 13027555 (SEQ ID NO:3) for the ACE gene.

Figure 4 depicts the amino acid sequence corresponding to reference GI 4503273 (SEQ ID NO:4) for the ACE protein.

Figure 5 depicts the nucleotide sequence corresponding to reference sequence GI 182597 (SEQ ID NO:5) for the FGB gene.

Figure 6 depicts the amino acid sequence corresponding to reference GI 11761631 (SEQ ID NO:6) for the FGB protein.

Detailed Description of the Invention

The present invention is based, in part, on the identification of polymorphic regions within the thrombospondin 2 (THBS2) gene, the angiotensin converting enzyme 1 (ACE) gene, and the beta fibrinogen (FGB) gene. The polymorphic regions of the invention contain polymorphisms which correlate with specific diseases or conditions, including vascular diseases or disorders, including, but not limited to, atherosclerosis, coronary artery disease (CAD), myocardial infarction (MI), ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

THBS2

Two SNPs in the THBS2 gene have been identified which are associated with vascular disease, *e.g.*, CAD and MI. The first THBS2 SNP, referred to herein as G5755e5, is a change from a thymidine (T) to a guanine (G) in the THBS2 gene at residue 3949 of the reference sequence GI 307505. The second THBS2 SNP, referred to herein as G5755e9, is a change from a thymidine (T) to a cytidine (C) in the THBS2 gene at residue 4476 of the reference sequence GI 307505. These SNPs are within the 3' untranslated region of the THBS2 gene. Therefore, they do not result in a change in the amino acid sequence of the THBS2 protein.

The variant allele, G, of the THBS2 SNP G5755e5, was previously shown to be associated with vascular disease, *e.g.*, MI and CAD. Individuals homozygous for the variant allele (GG) are at greater than 2-fold decreased odds of having vascular disease. Homozygous carriers of the variant allele of the G5755e9 SNP (CC) also showed a ~3-fold decreased odds of vascular disease.

These two SNPs, G5755e5 and G5755e9, are in significant negative linkage disequilibrium with each other (D'=.49 (-), p=.04). The two SNPs together reveal distinct patterns of risk. Pattern 1 comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT). Pattern 2 comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG). Patterns 1 and 2 may independently influence risk of vascular disease. Individuals who have pattern 1 or pattern 2 are at ~3-fold decreased

odds of vascular disease relative to persons with any other combination of genotypes for these two SNPs (odds ratio=0.32, p=.001). Thus, individuals with pattern 1 or pattern 2 are protected against vascular disease, *e.g.,* CAD and/or MI.

ACE

A SNP in the ACE gene, identified herein as G765u2, has been identified which is also associated with a decreased risk of vascular disease, *e.g.,* MI and CAD, in a subject. The G765u2 SNP is a change from an adenine (A) to a guanine (G) at nucleotide residue 86408 of the ACE reference sequence GI 13027555. This SNP is a "silent" variant. That is, it does not result in a change in the amino acid sequence of the ACE protein. Individuals with one copy of an A (the reference allele) and one copy of a G (the variant allele) at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype) are at a decreased risk for vascular disease, *e.g.,* CAD or MI (CAD odds ratio:0.71; MI odds ratio: 0.66) relative to persons with other genotypes for this SNP (*e.g.*, AA or GG genotypes).. Thus, individuals with this genotype are protected against vascular disease, *e.g.* CAD and/or MI.

An insertion/deletion polymorphism in the ACE gene was previously associated with vascular disease, *e.g.,* associated with a decreased risk for MI (as described in Cambien F, *et al.* (1992) *Nature* 359: 641-644, incorporated herein in its entirety by reference). The G765u2 SNP may be found to be in linkage disequilibrium with the previously identified insertion/deletion polymorphism. If these two polymorphisms are in linkage disequilibrium (LD), the G765u2 SNP would act as a marker for the insertion/deletion polymorphism. Regardless of LD between these two polymorphisms, the G765u2 SNP represents a novel association with vascular disease.

FGB

Two SNPs in the FGB gene, identified herein as FGBu1 and FGBu4, have been identified which are associated with a decreased risk of vascular disease, *e.g.,* CAD and/or MI. The first SNP, FGBu1, is a change from a cytidine (C) to a thymidine (T) at nucleotide residue 5118 of the FGB reference sequence GI 182597. This SNP is a silent variant. The second SNP, FGBu4, is a change from a guanine (G) to an adenine (A) at

nucleotide residue 8059 in the reference sequence GI 182597. This polymorphism is a missense variation which results in a change from an arginine (R) to a lysine (K) in the amino acid sequence of FGB (SEQ ID NO:6) at amino acid residue 478. For the FGBu1 SNP, individuals with two copies of a T (the variant allele) at nucleotide residue 5119 of the FGB reference sequence GI 182597 are at a decreased risk for vascular disease, *e.g.*, CAD or MI (CAD odds ratio: 0.28; MI odds ratio: 0.43) relative to persons with the CC genotype. Individuals with one copy of a T and one copy of a C (the reference allele) at nucleotide residue 5119 of the FGB reference sequence GI 182597 are also at a decreased risk for vascular disease, *e.g.*, CAD or MI (CAD odds ratio: 0.66; MI odds ratio: 0.72) relative to persons with the CC genotype. Thus, individuals with the TT or CT genotype at nucleotide residue 5119 of the FGB reference sequence GI 182597 are protected against vascular disease, *e.g.* CAD and/or MI.

For the FGBu4 SNP, individuals with two copies of an A (the variant allele) at nucleotide residue 8059 of the FGB reference sequence GI 182597 are at a decreased risk for vascular disease, *e.g.*, CAD or MI (CAD odds ratio: 0.28; MI odds ratio: 0.43) relative to persons with the GG genotype. Individuals with one copy of an A and one copy of a G (the reference allele) at nucleotide residue 5119 of the FGB reference sequence GI 182597 are also at a decreased risk for vascular disease, *e.g.*, CAD or MI (CAD odds ratio: 0.61; MI odds ratio: 0.66) relative to persons with the GG genotype. Thus, individuals with the AA or GA genotype at nucleotide residue 8059 of the FGB reference sequence GI 182597 are also protected against vascular disease, *e.g.* CAD and/or MI.

Other variants including one in the promoter region of the FGB gene at nucleotide residue –455 (as described in Shea S, *et al* (1999) *Am J Epidemiol*; 159:737-46, incorporated herein in its entirely by reference), have been previously associated with vascular disease, *e.g.*, CAD and MI. The FGBu1 and FGBu4 SNPs may be found to be in linkage disequilibrium with these previously identified SNPs. If these SNPs are in linkage disequilibrium (LD), the FGBu1 and FGBu4 SNPs would act as markers for the previously identified SNPs. Regardless of LD, the FGBu1 and FGBu4 SNPs represent novel associations with vascular disease.

The polymorphisms of the present invention are single nucleotide polymorphisms (SNPs) at a specific nucleotide residue within the THBS2 gene, the ACE gene, and FGB gene. The THBS2 gene, the ACE gene, and FGB gene have at least two alleles, referred to herein as the reference allele and the variant allele. The reference alleles (*i.e.*, the consensus sequences) have been designated based on their frequency in a general United States Caucasian population sample. The reference allele is the more common of the two alleles; the variant allele is the more rare of the two alleles. Nucleotide sequences in GenBank may correspond to either allele and correspond to the nucleotide sequence of the nucleotide sequence which has been deposited in GenBank™ and given a specific Accession Number (*e.g.*, GI 307505, the reference sequence for the THBS2 gene, GI 13027555, the reference sequence for the ACE gene, and GI 182597, the reference sequence for the FGB gene, corresponding to SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5, respectively). The reference sequence for the amino acid sequences of THBS2, ACE, and FGB proteins are set forth as SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6, respectively. The variant allele differs from the reference allele by at least one nucleotide at the site(s) identified in Tables 1, 4, and 6 (see Example 1, below), and those in linkage disequilibrium therewith. The present invention thus relates to nucleotides comprising variant alleles of the THBS2, ACE, and/or FGB reference sequences, and/or complements of the variant alleles to be used singly or in combination with each other.

The invention further relates to nucleotides comprising portions of the variant alleles and/or portions of complements of the variant alleles which comprise the site of the polymorphism and are at least 5 nucleotides or basepairs in length. Portions can be, for example, 5-10, 5-15, 10-20, 2-25, 10-30, 10-50 or 10-100 bases or basepairs long. For example, a portion of a variant allele which is 17 nucleotides or basepairs in length includes the polymorphism (*i.e.*, the nucleotide(s) which differ from the reference allele at that site) and twenty additional nucleotides or basepairs which flank the site in the variant allele. These additional nucleotides and basepairs can be on one or both sides of the polymorphism. Polymorphisms which are the subject of this invention are defined in Tables 1, 4, and 6 with respect to the reference sequences identified in Tables 1, 4, and 6 (GI 307505, GI 13027555, and GI 182597), and those polymorphisms in linkage

disequilibrium with the polymorphisms of Tables 1, 4, and 6. For example, the invention relates to nucleotides comprising a portion of the THBS2 gene having a nucleotide sequence of GI 307505 (SEQ ID NO:1), or a portion thereof, comprising a polymorphism at a specific nucleotide residue (*e.g.*, a guanine at nucleotide residue 3949 of GI 307505 or a cytidine at nucleotide residue 4476, or the complement thereof), nucleotides comprising a portion of the ACE gene having a nucleotide sequence of GI 13027555 (SEQ ID NO:3), or a portion thereof, comprising a polymorphism at a specific nucleotide residue (*e.g.*, a guanine at residue 86408, or the complement thereof), or nucleotides comprising a portion of the FGB gene having a nucleotide sequence of GI 182597 (SEQ ID NO:5), or a portion thereof, comprising a polymorphism at a specific nucleotide residue (*e.g.*, a thymidine at residue 5119 or an adenine at residue 8059, or the complement thereof).

Specific reference nucleotide (SEQ ID NO:1) and amino acid (SEQ ID NO: 2) sequences for THBS2 are shown in Figures 1 and 2, respectively. Specific reference nucleotide (SEQ ID NO:3) and amino acid (SEQ ID NO: 4) sequences for ACE are shown in Figures 3 and 4, respectively. Specific reference nucleotide (SEQ ID NO:5) and amino acid (SEQ ID NO: 6) sequences for FGB are shown in Figures 5 and 6, respectively. It is understood that the invention is not limited by these exemplified reference sequences, as variants of these sequences which differ at locations other than the SNP sites identified herein can also be utilized. The skilled artisan can readily determine the SNP sites in these other reference sequences which correspond to the SNP sites identified herein by aligning the sequence of interest with the reference sequences specifically disclosed herein. Programs for performing such alignments are commercially available. For example, the ALIGN program in the GCG software package can be used, utilizing a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4, for example.

The polymorphic region of the present invention is associated with specific diseases or disorders and has been identified in the human THBS2, ACE, and FGB genes by analyzing the DNA of human populations. In particular, 352 U.S. Caucasian gene by analyzing the DNA of cell lines derived from an ethnically diverse population by methods described in Cargill, *et al.* (1999) *Nature Genetics* 22:231-238.

Cases which were used to identify associations between vascular disease and SNPs were comprised of 352 U.S. Caucasian subjects with premature coronary artery disease were identified in 15 participating medical centers, fulfilling the criteria of either myocardial infarction, surgical or percutaneous revascularization, or a significant coronary artery lesion diagnosed before age 45 in men or age 50 in women and having a living sibling who met the same criteria. These cases were compared with a random sample of 418 Caucasian controls drawn from the general U.S. population in Atlanta, Georgia.

The allelic variants of the present invention were identified by performing denaturing high performance liquid chromatography (DHPLC) analysis, variant detector arrays (Affymetrix™), the polymerase chain reaction (PCR), and/or single stranded conformation polymorphism (SSCP) analysis of genomic DNA from independent individuals as described in the Examples, using PCR primers complementary to intronic sequences surrounding each of the exons, 3' UTR, and 5' upstream regulatory element sequences of the THBS2, ACE, and FGB genes.

The presence of at least one polymorphism in the ACE gene in the population studied was identified was identified and at least two polymorphisms in the THBS2, and FGB genes in the population studied were identified. Both of the variants are characterized as single nucleotide polymorphisms (SNPs). The preferred polymorphisms of the invention are listed in Tables 1, 4, and 6.

Tables 1, 4, and 6 contains a "polymorphism ID No." in column 2, which is used herein to identify each individual variant. In Tables 1, 4, and 6, the nucleotide sequence flanking each polymorphism is provided in column 9, wherein the polymorphic residue(s), having the variant nucleotide, is indicated in lower-case letters. There are 15 nucleotides flanking the polymorphic nucleotide residue (*i.e.*, 15 nucleotides 5' of the polymorphism and 15 nucleotides 3' of the polymorphism). Column 10 indicates the SEQ ID NO. that is used to identify each polymorphism. SEQ ID NOs.:7, 8, 9, 10, and 11 comprise sequences shown in column 9 with the variant nucleotide at the residue(s) shown in lower-case letters.

Each polymorphism is identified based on a change in the nucleotide sequence from a consensus sequence, or the "reference sequence." To identify the location of each polymorphism in Tables 1, 4, and 6, a specific nucleotide residue in a reference sequence is listed for each polymorphism, where nucleotide residue number 1 is the first (*i.e.*, 5') nucleotide in GI 307505 (the reference sequence for the THBS2 gene, corresponding to SEQ ID NO:1), the first nucleotide in GI 13027555 (the reference sequence for the ACE gene, corresponding to SEQ ID NO:3), and the first nucleotide in GI 182597 (the reference sequence for the FGB gene, corresponding to SEQ ID NO:5). Column 8 lists the reference sequence and polymorphic residue for each polymorphism.

Column 4 describes the type of variant for each SNP. The SNPs of the present invention result in either a silent variant, a missense variant, or a 3' untranslated region variant. For example, as can be seen in Tables 1, 4, and 6, both THBS2 SNPs (G5755e5 and G5755e9) are located in the 3' UTR of the THBS2 gene. The ACE SNP (G765u2) is a silent variant. The FGBu1 SNP in the FGB gene is also a silent variant. The FGBu4 SNP in the FGB gene results in a change from an arginine (R) to a lysine (K). Therefore, this SNP is identified as a missense SNP.

The nucleic acid molecules of the invention can be double- or single-stranded. Accordingly, the invention further provides for the complementary nucleic acid strands comprising the polymorphisms listed in Tables 1, 4, and 6.

The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. Such oligonucleotides will hybridize to one polymorphic form of the nucleic acid molecules described herein but not to the other polymorphic form(s) of the sequence. Thus such oligonucleotides can be used to determine the presence or absence of particular alleles of the polymorphic sequences described herein. These oligonucleotides can be probes or primers.

Not only does the present invention provide polymorphisms in linkage disequilibrium with the polymorphisms of Tables 1, 4, and 6, it also provides methods for revealing the existence of yet other polymorphic regions in the human THBS2, ACE, or FGB gene. For example, the polymorphism studies described herein can also be applied to populations in which other vascular diseases or disorders are prevalent.

Other aspects of the invention are described below or will be apparent to one of skill in the art in light of the present disclosure.

Definitions

For convenience, the meaning of certain terms and phrases employed in the specification, examples, and appended claims are provided below.

The term "allele", which is used interchangeably herein with "allelic variant" refers to alternative forms of a gene or portions thereof. Alleles occupy the same locus or position on homologous chromosomes. When a subject has two identical alleles of a gene, the subject is said to be homozygous for the gene or allele. When a subject has two different alleles of a gene, the subject is said to be heterozygous for the gene or allele. Alleles of a specific gene, including the THBS2, ACE, or FGB genes, can differ from each other in a single nucleotide, or several nucleotides, and can include substitutions, deletions, and insertions of nucleotides. An allele of a gene can also be a form of a gene containing one or more mutations.

The term "allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene" refers to an alternative form of the THBS2, ACE, or FGB gene having one of several possible nucleotide sequences found in that region of the gene in the population.

"Biological activity" or "bioactivity" or "activity" or "biological function", which are used interchangeably, for the purposes herein when applied to THBS2, ACE, or FGB, means an effector or antigenic function that is directly or indirectly performed by a THBS2, ACE, or FGB polypeptide (whether in its native or denatured conformation), or by a fragment thereof. Biological activities include modulation of the development of atherosclerotic plaque leading to vascular disease and other biological activities, whether presently known or inherent. A THBS2, ACE, or FGB bioactivity can be modulated by directly affecting a THBS2, ACE, or FGB protein effected by, for example, changing the level of effector or substrate level. Alternatively, a THBS2, ACE, or FGB bioactivity can be modulated by modulating the level of a THBS2, ACE, or FGB protein, such as by modulating expression of a THBS2, ACE, or FGB gene. Antigenic functions include possession of an epitope or antigenic site that is capable of

cross-reacting with antibodies that bind a native or denatured THBS2, ACE, or FGB polypeptide or fragment thereof.

Biologically active THBS2, ACE, or FGB polypeptides include polypeptides having both an effector and antigenic function, or only one of such functions. THBS2, ACE, or FGB polypeptides include antagonist polypeptides and native THBS2, ACE, or FGB polypeptides, provided that such antagonists include an epitope of a native THBS2, ACE, or FGB polypeptide. An effector function of THBS2, ACE, or FGB polypeptide can be the ability to bind to a ligand of a THBS2, ACE, or FGB molecule.

As used herein the term "bioactive fragment of a THBS2, ACE, or FGB protein" refers to a fragment of a full-length THBS2, ACE, or FGB protein, wherein the fragment specifically mimics or antagonizes the activity of a wild-type THBS2, ACE, or FGB protein. The bioactive fragment preferably is a fragment capable of binding to a second molecule, such as a ligand.

The term "an aberrant activity" or "abnormal activity", as applied to an activity of a protein such as THBS2, ACE, or FGB, refers to an activity which differs from the activity of the wild-type (*i.e.*, normal or reference) protein or which differs from the activity of the protein in a healthy subject, *e.g.*, a subject not afflicted with a disease associated with a THBS2, ACE, or FGB allelic variant. An activity of a protein can be aberrant because it is stronger than the activity of its wild-type counterpart. Alternatively, an activity of a protein can be aberrant because it is weaker or absent relative to the activity of its wild-type counterpart. An aberrant activity can also be a change in reactivity. For example an aberrant protein can interact with a different protein or ligand relative to its wild-type counterpart. A cell can also have aberrant THBS2, ACE, or FGB activity due to overexpression or underexpression of the THBS2, ACE, or FGB gene. Aberrant THBS2, ACE, or FGB activity can result from a mutation in the gene, which results, *e.g.*, in lower or higher binding affinity of a ligand to the THBS2, ACE, or FGB protein encoded by the mutated gene. Aberrant THBS2, ACE, or FGB activity can also result from an abnormal THBS2, ACE, or FGB 5' upstream regulatory element activity.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular cell but to the progeny or derivatives of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

As used herein, the term "course of clinical therapy" refers to any chosen method to treat, prevent, or ameliorate a vascular disease, *e.g.,* CAD or MI, symptoms thereof, or related diseases or disorders. Courses of clinical therapy include, but are not limited to, lifestyle changes (*e.g.,* changes in diet or environment), administration of medication, use of medical devices, such as, but not limited to, stents, angioplasty devices, defibrillators, pacemakers, and surgical procedures, such as, for example, percutaneous transluminal coronary balloon angioplasty (PTCA) or laser angioplasty, defibrillators, implantation of a stent, or other surgical intervention, such as, for example, coronary bypass grafting (CABG), or any combination thereof.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid molecule comprising an open reading frame and including at least one exon and (optionally) an intron sequence. The term "intron" refers to a DNA sequence present in a given gene which is spliced out during mRNA maturation.

As used herein, the term "genetic profile" refers to the information obtained from identification of the specific alleles of a subject, *e.g.,* specific alleles within a polymorphic region of a particular gene or genes or proteins encoded by such genes. For example, a THBS genetic profile refers to the specific alleles of a subject within the THBS2 gene, an ACE genetic profile refers to the specific alleles of a subject within the ACE gene, and a FGB genetic profile refers to the specific alleles of a subject within the FGB gene. For example, one can determine a subject's THBS2, ACE, and/or FGB genetic profile by determining the identity of the nucleotide present at nucleotide position 3949 and/or nucleotide position 4476 of SEQ ID NO:1, and/or the nucleotide present at nucleotide position 86408 of SEQ ID NO:3, and/or the nucleotide present at nucleotide position 5119 and/or nucleotide position 8059 of SEQ ID NO:5. One can also determine a subject's FGB genetic profile by determining the identity of the amino

acid present at amino acid residue 478 of SEQ ID NO:6. The genetic profile of a particular disease can be ascertained through identification of the identity of allelic variants in one or more genes which are associated with the particular disease.

"Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 % identity, though preferably less than 25 % identity, with one of the sequences of the present invention.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = number of identical positions/total number of positions (*e.g.*, overlapping positions) x100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences

homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Yet another useful algorithm for identifying regions of local sequence similarity and alignment is the FASTA algorithm as described in Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85:2444-2448. When using the FASTA algorithm for comparing nucleotide or amino acid sequences, a PAM120 weight residue table can, for example, be used with a *k*-tuple value of 2.

The term "a homolog of a nucleic acid" refers to a nucleic acid having a nucleotide sequence having a certain degree of homology with the nucleotide sequence of the nucleic acid or complement thereof. For example, a homolog of a double stranded nucleic acid having SEQ ID NO:N is intended to include nucleic acids having a nucleotide sequence which has a certain degree of homology with SEQ ID NO:N or with the complement thereof. Preferred homologs of nucleic acids are capable of hybridizing to the nucleic acid or complement thereof.

The term "hybridization probe" or "primer" as used herein is intended to include oligonucleotides which hybridize bind in a base-specific manner to a complementary strand of a target nucleic acid. Such probes include peptide nucleic acids, and described in Nielsen *et al.*, (1991) *Science* 254:1497-1500. Probes and primers can be any length suitable for specific hybridization to the target nucleic acid sequence. The most appropriate length of the probe and primer may vary depending on the hybridization

method in which it is being used; for example, particular lengths may be more appropriate for use in microfabricated arrays, while other lengths may be more suitable for use in classical hybridization methods. Such optimizations are known to the skilled artisan. Suitable probes and primers can range form about 5 nucleotides to about 30 nucleotides in length. For example, probes and primers can be 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28 or 30 nucleotides in length. The probe or primer of the invention comprises a sequence that flanks and/or preferably overlaps, at least one polymorphic site occupied by any of the possible variant nucleotides. The nucleotide sequence of an overlapping probe or primer can correspond to the coding sequence of the allele or to the complement of the coding sequence of the allele.

The term "vascular disease or disorder" as used herein refers to any disease or disorder effecting the vascular system, including the heart and blood vessels. A vascular disease or disorder includes any disease or disorder characterized by vascular dysfunction, including, for example, intravascular stenosis (narrowing) or occlusion (blockage), due to the development of atherosclerotic plaque and diseases and disorders resulting therefrom. Examples of vascular diseases and disorders include, without limitation, atherosclerosis, CAD, MI, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

The term "interact" as used herein is meant to include detectable interactions between molecules, such as can be detected using, for example, a binding or hybridization assay. The term interact is also meant to include "binding" interactions between molecules. Interactions may be, for example, protein-protein, protein-nucleic acid, protein-small molecule or small molecule-nucleic acid in nature.

The term "intronic sequence" or "intronic nucleotide sequence" refers to the nucleotide sequence of an intron or portion thereof.

The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs or RNAs, respectively, that are present in the natural source of the macromolecule. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an

"isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides.

The term "linkage" describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome. It can be measured by percent recombination between the two genes, alleles, loci, or genetic markers. The term "linkage disequilibrium" refers to a greater than random association between specific alleles at two marker loci within a particular population. In general, linkage disequilibrium decreases with an increase in physical distance. If linkage disequilibrium exists between two markers, then the genotypic information at one marker can be used to make probabilistic predictions about the genotype of the second marker.

The term "locus" refers to a specific position in a chromosome. For example, a locus of a THBS2, ACE, or FGB gene refers to the chromosomal position of the THBS2, ACE, or FGB gene.

The term "modulation" as used herein refers to both upregulation, (*i.e.*, activation or stimulation), for example by agonizing; and downregulation (i.e. inhibition or suppression), for example by antagonizing of a bioactivity (*e.g.* expression of a gene).

The term "molecular structure" of a gene or a portion thereof refers to the structure as defined by the nucleotide content (including deletions, substitutions, additions of one or more nucleotides), the nucleotide sequence, the state of methylation, and/or any other modification of the gene or portion thereof.

The term "mutated gene" refers to an allelic form of a gene that differs from the predominant form in a population. A mutated gene is capable of altering the phenotype of a subject having the mutated gene relative to a subject having the predominant form of the gene. If a subject must be homozygous for this mutation to have an altered phenotype, the mutation is said to be recessive. If one copy of the mutated gene is sufficient to alter the phenotype of the subject, the mutation is said to be dominant. If a subject has one copy of the mutated gene and has a phenotype that is intermediate

between that of a homozygous and that of a heterozygous subject (for that gene), the mutation is said to be co-dominant.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, derivatives, variants and analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine, and deoxythymidine. For purposes of clarity, when referring herein to a nucleotide of a nucleic acid, which can be DNA or an RNA, the terms "adenine", "cytidine", "guanine", and thymidine" and/or "A", "C", "G", and "T", respectively, are used. It is understood that if the nucleic acid is RNA, a nucleotide having a uracil base is uridine.

The term "nucleotide sequence complementary to the nucleotide sequence set forth in SEQ ID NO:N" refers to the nucleotide sequence of the complementary strand of a nucleic acid strand having SEQ ID NO:N. The term "complementary strand" is used herein interchangeably with the term "complement". The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand. When referring to double stranded nucleic acids, the complement of a nucleic acid having SEQ ID NO:N refers to the complementary strand of the strand having SEQ ID NO:N or to any nucleic acid having the nucleotide sequence of the complementary strand of SEQ ID NO:N. When referring to a single stranded nucleic acid having the nucleotide sequence SEQ ID NO:N, the complement of this nucleic acid is a nucleic acid having a nucleotide sequence which is complementary to that of SEQ ID NO:N. The nucleotide sequences and complementary sequences thereof are always given in the 5' to 3' direction. The term "complement" and "reverse complement" are used interchangeably herein.

A "non-human animal" of the invention can include mammals such as rodents, non-human primates, sheep, goats, horses, dogs, cows, chickens, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians, such as members of the

Xenopus genus, and transgenic chickens can also provide important tools for understanding and identifying agents which can affect, for example, embryogenesis and tissue formation. The term "chimeric animal" is used herein to refer to animals in which an exogenous sequence is found, or in which an exogenous sequence is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that an exogenous sequence is present and/or expressed or disrupted in some tissues, but not others.

The term "oligonucleotide" is intended to include and single- or double stranded DNA or RNA. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of THBS2, ACE, or FGB gene sequence or their complements, which include and/or flank any one of the polymorphic sites shown in Tables 1, 4, and 6. The segments can be between 5 and 250 bases, and, in specific embodiments, are between 5-10, 5-20, 10-20, 10-50, 20-50 or 10-100 bases. For example, the segments can be 21 bases. The polymorphic site can occur within any position of the segment or a region next to the segment. The segments can be from any of the allelic forms of THBS2, ACE, or FGB gene sequence shown in Tables 1, 4, and 6.

The term "operably-linked" is intended to mean that the 5' upstream regulatory element is associated with a nucleic acid in such a manner as to facilitate transcription of the nucleic acid from the 5' upstream regulatory element.

The term "polymorphism" refers to the coexistence of more than one form of a gene or portion thereof. A portion of a gene of which there are at least two different forms, *i.e.*, two different nucleotide sequences, is referred to as a "polymorphic region of a gene." A polymorphic locus can be a single nucleotide, the identity of which differs in the other alleles. A polymorphic locus can also be more than one nucleotide long. The allelic form occurring most frequently in a selected population is often referred to as the reference and/or wildtype form. Other allelic forms are typically designated or alternative or variant alleles. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic or biallelic polymorphism has two forms. A trialleleic polymorphism has three forms.

A "polymorphic gene" refers to a gene having at least one polymorphic region.

The term "primer" as used herein, refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and as agent for polymerization, such as DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The length of a primer may vary but typically ranges from 15 to 30 nucleotides. A primer need not match the exact sequence of a template, but must be sufficiently complementary to hybridize with the template.

The term "primer pair" refers to a set of primers including an upstream primer that hybridizes with the 3' end of the complement of the DNA sequence to be amplified and a downstream primer that hybridizes with the 3' end of the sequence to be amplified.

The terms "protein", "polypeptide" and "peptide" are used interchangeably herein when referring to a gene product.

The term "recombinant protein" refers to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein.

A "regulatory element", also termed herein "regulatory sequence" is intended to include elements which are capable of modulating transcription from a 5' upstream regulatory sequence, including, but not limited to a basic promoter, and include elements such as enhancers and silencers. The term "enhancer", also referred to herein as "enhancer element", is intended to include regulatory elements capable of increasing, stimulating, or enhancing transcription from a 5' upstream regulatory element, including a basic promoter. The term "silencer", also referred to herein as "silencer element" is intended to include regulatory elements capable of decreasing, inhibiting, or repressing transcription from a 5' upstream regulatory element, including a basic promoter. Regulatory elements are typically present in 5' flanking regions of genes. Regulatory elements also may be present in other regions of a gene, such as introns. Thus, it is possible that THBS2, ACE, or FGB genes have regulatory elements located in introns, exons, coding regions, and 3' flanking sequences. Such regulatory elements are also

intended to be encompassed by the present invention and can be identified by any of the assays that can be used to identify regulatory elements in 5' flanking regions of genes.

The term "regulatory element" further encompasses "tissue specific" regulatory elements, *i.e.,* regulatory elements which effect expression of an operably linked DNA sequence preferentially in specific cells (*e.g.,* cells of a specific tissue). Gene expression occurs preferentially in a specific cell if expression in this cell type is significantly higher than expression in other cell types. The term "regulatory element" also encompasses non-tissue specific regulatory elements, *i.e.,* regulatory elements which are active in most cell types. Furthermore, a regulatory element can be a constitutive regulatory element, *i.e.,* a regulatory element which constitutively regulates transcription, as opposed to a regulatory element which is inducible, *i.e.,* a regulatory element which is active primarily in response to a stimulus. A stimulus can be, *e.g.,* a molecule, such as a protein, hormone, cytokine, heavy metal, phorbol ester, cyclic AMP (cAMP), or retinoic acid.

Regulatory elements are typically bound by proteins, *e.g.,* transcription factors. The term "transcription factor" is intended to include proteins or modified forms thereof, which interact preferentially with specific nucleic acid sequences, *i.e.,* regulatory elements, and which in appropriate conditions stimulate or repress transcription. Some transcription factors are active when they are in the form of a monomer. Alternatively, other transcription factors are active in the form of a dimer consisting of two identical proteins or different proteins (heterodimer). Modified forms of transcription factors are intended to refer to transcription factors having a postranslational modification, such as the attachment of a phosphate group. The activity of a transcription factor is frequently modulated by a postranslational modification. For example, certain transcription factors are active only if they are phosphorylated on specific residues. Alternatively, transcription factors can be active in the absence of phosphorylated residues and become inactivated by phosphorylation. A list of known transcription factors and their DNA binding site can be found, *e.g.,* in public databases, *e.g.,* TFMATRIX Transcription Factor Binding Site Profile database.

The term "single nucleotide polymorphism" (SNP) refers to a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (*e.g.*, sequences that vary in less than 1/100 or 1/1000 members of a population). A SNP usually arises due to substitution of one nucleotide for another at the polymorphic site. SNPs can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele. Typically the polymorphic site is occupied by a base other than the reference base. For example, where the reference allele contains the base "T" (thymidine) at the polymorphic site, the altered allele can contain a "C" (cytidine), "G" (guanine), or "A" (adenine) at the polymorphic site.

SNP's may occur in protein-coding nucleic acid sequences, in which case they may give rise to a defective or otherwise variant protein, or genetic disease. Such a SNP may alter the coding sequence of the gene and therefore specify another amino acid (a "missense" SNP) or a SNP may introduce a stop codon (a "nonsense" SNP). When a SNP does not alter the amino acid sequence of a protein, the SNP is called "silent." SNP's may also occur in noncoding regions of the nucleotide sequence. This may result in defective protein expression, *e.g.*, as a result of alternative spicing, or it may have no effect.

As used herein, the term "specifically hybridizes" or "specifically detects" refers to the ability of a nucleic acid molecule of the invention to hybridize to at least approximately 6, 8, 10, 12, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130 or 140 consecutive nucleotides of either strand of a THBS2, ACE, or FGB gene.

As used herein, the term "transfection" means the introduction of a nucleic acid, *e.g.*, an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. The term "transduction" is generally used herein when the transfection with a nucleic acid is by viral delivery of the nucleic acid. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a polypeptide or, in the case of anti-sense expression from the transferred gene, the expression of a naturally-occurring form of the recombinant protein is disrupted.

As used herein, the term "transgene" refers to a nucleic acid sequence which has been genetic-engineered into a cell. Daughter cells deriving from a cell in which a transgene has been introduced are also said to contain the transgene (unless it has been deleted). A transgene can encode, *e.g.*, a polypeptide, or an antisense transcript, partly or entirely heterologous, *i.e.*, foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (*e.g.*, it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). Alternatively, a transgene can also be present in an episome. A transgene can include one or more transcriptional regulatory sequence and any other nucleic acid, (*e.g.* intron), that may be necessary for optimal expression of a selected nucleic acid.

A "transgenic animal" refers to any animal, preferably a non-human animal, *e.g.* a mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by genetic engineering, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of one of a protein, *e.g.* either agonistic or antagonistic forms. However, transgenic animals in which the recombinant gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more genes is caused by human intervention, including both recombination and antisense techniques.

The term "treatment", or "treating" as used herein, is defined as the application or administration of a therapeutic agent to a subject, implementation of lifestyle changes (*e.g.*, changes in diet or environment), administration of medication, use of medical devices, such as, but not limited to, stents, angioplasty devices, defibrillators, and surgical procedures, such as, for example, percutaneous transluminal coronary balloon angioplasty (PTCA) or laser angioplasty, implantation of a stent, or other surgical intervention, such as, for example, coronary bypass grafting (CABG), or any combination thereof, or application or administration of a therapeutic agent to an isolated tissue or cell line from a subject, who has a disease or disorder, a symptom of disease or disorder or a predisposition toward a disease or disorder, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease or disorder, the symptoms of the disease or disorder, or the predisposition toward disease.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting or replicating another nucleic acid to which it has been linked. One type of preferred vector is an episome, *i.e.*, a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively-linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA circles which, in their vector form are not physically linked to the host chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

Polymorphisms Used in the Methods of the Invention

The nucleic acid molecules of the present invention include specific allelic variants of the THBS2, ACE, and FGB genes, which differ from the reference sequences set forth in SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5, respectively, or at least a portion thereof, having a polymorphic region. The preferred nucleic acid molecules of

the present invention comprise THBS2, ACE, and FGB sequences having one or more of the polymorphisms shown in Tables 1, 4, and 6 (SEQ ID NOs.:7, 8, 9, 10, and 11), and those in linkage disequilibrium therewith. The invention further comprises isolated nucleic acid molecules complementary to nucleic acid molecules comprising the polymorphisms of the present invention. Nucleic acid molecules of the present invention can function as probes or primers, *e.g.*, in methods for determining the allelic identity of a THBS2, ACE, or FGB polymorphic region. The nucleic acids of the invention can also be used, singly, or in combination, to determine whether a subject is or is not at risk of developing a disease associated with a specific allelic variant of a THBS2, ACE, or FGB polymorphic region, *e.g.*, a vascular disease or disorder. The nucleic acids of the invention can further be used to prepare or express THBS2, ACE, or FGB polypeptides encoded by specific alleles, such as mutant alleles. Such nucleic acids can be used in gene therapy. Polypeptides encoded by specific THBS2, ACE, or FGB alleles, such as mutant THBS2, ACE, or FGB polypeptides, can also be used in therapy or for preparing reagents, *e.g.*, antibodies, for detecting THBS2, ACE, or FGB proteins encoded by these alleles. Accordingly, such reagents can be used to detect mutant THBS2, ACE, or FGB proteins.

As described herein, allelic variants of human THBS2, ACE, or FGB genes have been identified. The invention is intended to encompass these allelic variants as well as, those in linkage disequilibrium which can be identified, *e.g.*, according to the methods described herein. "Linkage disequilibrium" refers to an association between specific alleles at two marker loci within a particular population. In general, linkage disequilbrium decreases with an increase in physical distance. If linkage disequilbrium exists between two markers, then the genotypic information at one marker can be used to make predictions about the genotype of the second marker.

The invention also provides isolated nucleic acids comprising at least one polymorphic region of a THBS2, ACE, or FGB gene having a nucleotide sequence which differs from the reference nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5 respectively. Preferred nucleic acids have a variant allele located in the coding region of a THBS2, ACE, or FGB gene, the upstream regulatory element, an exon, or in the 3' UTR of a THBS2, ACE, or FGB gene. Accordingly,

preferred nucleic acids of the invention comprise a guanine at residue 3949 of GI 307505, and/or a cytidine at residue 4476 of GI 307505 (as set forth in SEQ ID NO:1), or the complement thereof, and/or a guanine at residue 455299 of GI 13027555 (as set forth in SEQ ID NO:3), or the complement thereof, and/or a thymidine at residue 5119 of GI 182597, and/or an adenine at residue 8059 of GI 182597 (set forth herein as SEQ ID NO:5). Preferred nucleic acids used in combination in the methods of the invention to predict decreased risk of vascular diseases or disorders comprise "pattern 1," which comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT) or "pattern 2", which comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG) is at approximately 3-fold decreased odds of vascular disease.

Other preferred nucleic acids used in the methods of the invention to predict decreased risk of vascular diseases or disorders comprise one copy of an A and one copy of a G at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype) is at a decreased risk for vascular disease.

Still other preferred nucleic acids used in the methods of the invention to predict decreased risk of vascular diseases or disorders comprise two copies of a T at nucleotide residue 5119 of the FGB reference sequence GI 182597 is at a decreased risk for vascular disease, *e.g.*, CAD and MI. A subject having one copy of a T and one copy of a C at nucleotide residue 5119 of the FGB reference sequence GI 182597 is also at a decreased risk for vascular disease, *e.g.*, CAD and MI.

Other preferred nucleic acids used in the methods of the invention to predict decreased risk of vascular diseases or disorders comprise two copies of an A at nucleotide residue 8059 of the FGB reference sequence GI 182597 is at a decreased risk for vascular disease. A subject having one copy of an A and one copy of a G at nucleotide residue 5119 of the FGB reference sequence GI 182597 is also at a decreased risk for vascular disease (see Example 1, below).

The nucleic acid molecules of the present invention can be single stranded DNA (*e.g.*, an oligonucleotide), double stranded DNA (*e.g.*, double stranded oligonucleotide) or RNA. Preferred nucleic acid molecules of the invention can be used as probes or primers. Primers of the invention refer to nucleic acids which hybridize to a nucleic acid

sequence which is adjacent to the region of interest or which covers the region of interest and is extended. As used herein, the term "hybridizes" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions vary according to the length of the involved nucleotide sequence but are known to those skilled in the art and can be found or determined based on teachings in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional stringent conditions and formulas for determining such conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions for hybrids that are at least basepairs in length includes hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions for such hybrids includes hybridization in 1X SSC, at about 65-70°C (or hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions for such hybrids includes hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45°C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70° C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete.

The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}$[Na^+]) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and [Na^+] is the concentration of sodium ions in the hybridization buffer ([Na^+] for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH_2PO_4, 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH_2PO_4, 1% SDS at 65°C, see *e.g.*, Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995, (or alternatively 0.2X SSC, 1% SDS).

A primer or probe can be used alone in a detection method, or a primer can be used together with at least one other primer or probe in a detection method. Primers can also be used to amplify at least a portion of a nucleic acid. Probes of the invention refer to nucleic acids which hybridize to the region of interest and which are not further extended. For example, a probe is a nucleic acid which specifically hybridizes to a polymorphic region of a THBS2, ACE, or FGB gene, and which by hybridization or absence of hybridization to the DNA of a subject or the type of hybrid formed will be indicative of the identity of the allelic variant of the polymorphic region of the THBS2, ACE, or FGB gene.

Numerous procedures for determining the nucleotide sequence of a nucleic acid molecule, or for determining the presence of mutations in nucleic acid molecules include a nucleic acid amplification step, which can be carried out by, *e.g.*, polymerase chain reaction (PCR). Accordingly, in one embodiment, the invention provides primers for amplifying portions of a THBS2, ACE, or FGB gene, such as portions of exons and/or

portions of introns. In a preferred embodiment, the exons and/or sequences adjacent to the exons of the human THBS2, ACE, or FGB gene will be amplified to, *e.g.*, detect which allelic variant, if any, of a polymorphic region is present in the THBS2, ACE, or FGB gene of a subject. Preferred primers comprise a nucleotide sequence complementary a specific allelic variant of a THBS2, ACE, or FGB polymorphic region and of sufficient length to selectively hybridize with a THBS2, ACE, or FGB gene. In a preferred embodiment, the primer, *e.g.*, a substantially purified oligonucleotide, comprises a region having a nucleotide sequence which hybridizes under stringent conditions to about 6, 8, 10, or 12, preferably 25, 30, 40, 50, or 75 consecutive nucleotides of a THBS2, ACE, or FGB gene. In an even more preferred embodiment, the primer is capable of hybridizing to a THBS2, ACE, or FGB nucleotide sequence, complements thereof, allelic variants thereof, or complements of allelic variants thereof. For example, primers comprising a nucleotide sequence of at least about 8, 10, 12, or 15 consecutive nucleotides, at least about 25 nucleotides or having from about 15 to about 20 nucleotides set forth in any of SEQ ID NOs:7, 8, 9, 10, or 11, or complement thereof are provided by the invention. Primers having a sequence of more than about 25 nucleotides are also within the scope of the invention. Preferred primers of the invention are primers that can be used in PCR for amplifying each of the exons of a THBS2, ACE, or FGB gene.

Primers can be complementary to nucleotide sequences located close to each other or further apart, depending on the use of the amplified DNA. For example, primers can be chosen such that they amplify DNA fragments of at least about 10 nucleotides or as much as several kilobases. Preferably, the primers of the invention will hybridize selectively to THBS2, ACE, or FGB nucleotide sequences located about 150 to about 350 nucleotides apart.

For amplifying at least a portion of a nucleic acid, a forward primer (*i.e.*, 5' primer) and a reverse primer (*i.e.*, 3' primer) will preferably be used. Forward and reverse primers hybridize to complementary strands of a double stranded nucleic acid, such that upon extension from each primer, a double stranded nucleic acid is amplified. A forward primer can be a primer having a nucleotide sequence or a portion of the nucleotide sequence shown in Tables 1, 4, and 6 (*e.g.*, SEQ ID NOs.:7, 8, 9, 10, and 11).

A reverse primer can be a primer having a nucleotide sequence or a portion of the nucleotide sequence that is complementary to a nucleotide sequence shown in Tables 1, 4, and 6 (*e.g.*, SEQ ID NOs.:7, 8, 9, 10, and 11).

Yet other preferred primers of the invention are nucleic acids which are capable of selectively hybridizing to an allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene. Thus, such primers can be specific for a THBS2, ACE, or FGB gene sequence, so long as they have a nucleotide sequence which is capable of hybridizing to a THBS2, ACE, or FGB gene. Preferred primers are capable of specifically hybridizing to any of the allelic variants listed in Tables 1, 4, and 6. Such primers can be used, *e.g.*, in sequence specific oligonucleotide priming as described further herein.

Other preferred primers used in the methods of the invention are nucleic acids which are capable of hybridizing to the reference sequence of a THBS2, ACE, or FGB gene, thereby detecting the presence of the reference allele of an allelic variant or the absence of a variant allele in the THBS2, ACE, or FGB genes and primers capable of hybridizing to the variant sequence of a THBS2, ACE, or FGB gene. Such primers can be used in combination, *e.g.*, primers specific for the alleles of pattern 1 or pattern 2, as described herein. The sequences of primers specific for the reference sequences comprising the THBS2, ACE, or FGB genes will be readily apparent to one of skill in the art.

The THBS2, ACE, or FGB nucleic acids of the invention can also be used as probes, *e.g.*, in therapeutic and diagnostic assays. For instance, the present invention provides a probe comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region having a nucleotide sequence that is capable of hybridizing specifically to a region of a THBS2, ACE, or FGB gene which is polymorphic (*e.g.*, SEQ ID NOs.:7, 8, 9, 10, and 11, or a portion thereof). In an even more preferred embodiment of the invention, the probes are capable of hybridizing specifically to one allelic variant of a THBS2, ACE, or FGB gene having a nucleotide sequence which differs from the nucleotide sequence set forth in SEQ ID NOs: 1, 3, or 5. Such probes can then be used to specifically detect which allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene is present in a subject. The

polymorphic region can be located in the 5' upstream regulatory element, exon, or intron sequences of a THBS2, ACE, or FGB gene.

Particularly, preferred probes of the invention have a number of nucleotides sufficient to allow specific hybridization to the target nucleotide sequence. Where the target nucleotide sequence is present in a large fragment of DNA, such as a genomic DNA fragment of several tens or hundreds of kilobases, the size of the probe may have to be longer to provide sufficiently specific hybridization, as compared to a probe which is used to detect a target sequence which is present in a shorter fragment of DNA. For example, in some diagnostic methods, a portion of a THBS2, ACE, or FGB gene may first be amplified and thus isolated from the rest of the chromosomal DNA and then hybridized to a probe. In such a situation, a shorter probe will likely provide sufficient specificity of hybridization. For example, a probe having a nucleotide sequence of about 10 nucleotides may be sufficient.

In preferred embodiments, the probe or primer further comprises a label attached thereto, which, *e.g.,* is capable of being detected, *e.g.* the label group is selected from amongst radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors.

In a preferred embodiment of the invention, the isolated nucleic acid, which is used, *e.g.,* as a probe or a primer, is modified, so as to be more stable than naturally occurring nucleotides. Exemplary nucleic acid molecules which are modified include phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patent Numbers 5,176,996; 5,264,564; and 5,256,775).

The nucleic acids of the invention can also be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule. The nucleic acids, *e.g.,* probes or primers, may include other appended groups such as peptides (*e.g.,* for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.,* Letsinger *et al.*, (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84:648-652; PCT Publication No. WO88/09810, published December 15, 1988), hybridization-triggered cleavage agents. (See, *e.g.,* Krol *et al.*, (1988) *BioTechniques* 6:958-976) or intercalating agents (See, *e.g.,* Zon, (1988) *Pharm. Res.* 5:539-549). To this end, the nucleic acid of the invention may be conjugated to another molecule, *e.g.,* a peptide,

hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The isolated nucleic acid comprising a THBS2, ACE, or FGB intronic sequence may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytidine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytidine, 5-methylcytidine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytidine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

The isolated nucleic acid may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the nucleic acid comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet a further embodiment, the nucleic acid is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier *et al.*, 1987, *Nucl. Acids Res.* 15:6625-6641). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue *et al.*, (1987) *Nucl. Acids Res.* 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.*, (1987) *FEBS Lett.* 215:327-330).

Any nucleic acid fragment of the invention can be prepared according to methods well known in the art and described, *e.g.,* in Sambrook, J. Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. For example, discrete fragments of the DNA can be prepared and cloned using restriction enzymes. Alternatively, discrete fragments can be prepared using the Polymerase Chain Reaction (PCR) using primers having an appropriate sequence.

Oligonucleotides of the invention may be synthesized by standard methods known in the art, *e.g.* by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein *et al.* ((1988) *Nucl. Acids Res.* 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, (1988), *Proc. Natl. Acad. Sci. U.S.A.* 85:7448-7451), etc.

The invention also provides vectors and plasmids comprising the nucleic acids of the invention. For example, in one embodiment, the invention provides a vector comprising at least a portion of the THBS2, ACE, or FGB gene comprising a polymorphic region. Thus, the invention provides vectors for expressing at least a portion of the newly identified allelic variants of the human THBS2, ACE, or FGB gene, as well as other allelic variants, comprising a nucleotide sequence which is different from the nucleotide sequence disclosed in GI 307505, GI 13027555, or GI 182597, respectively. The allelic variants can be expressed in eukaryotic cells, *e.g.,* cells of a subject, or in prokaryotic cells.

In one embodiment, the vector comprising at least a portion of a THBS2, ACE, or FGB allele is introduced into a host cell, such that a protein encoded by the allele is synthesized. The THBS2, ACE, or FGB protein produced can be used, *e.g.,* for the production of antibodies, which can be used, *e.g.,* in methods for detecting mutant forms of THBS2, ACE, or FGB. Alternatively, the vector can be used for gene therapy, and be, *e.g.,* introduced into a subject to produce THBS2, ACE, or FGB protein. Host cells comprising a vector having at least a portion of a THBS2, ACE, or FGB gene are also within the scope of the invention.

Polypeptides of the invention

The present invention provides isolated THBS2, ACE, or FGB polypeptides, such as THBS2, ACE, or FGB polypeptides which are encoded by specific allelic variants of THBS2, ACE, or FGB, including those identified herein, *e.g.*, proteins encoded by nucleic acids which differ from the reference sequence of THBS2, ACE, or FGB, or a portion thereof, as set forth herein. The amino acid sequences of the THBS2, ACE, or FGB proteins have been deduced. The THBS2 gene encodes a 1,172 amino acid protein and is described in, for example, LaBelle, *et al.* (1993) *Genomics* 17(1):225. The ACE gene encodes a 1,306 amino acid protein and is described in, for example, Rieder M.J. *et al.* (1999) *Nature Genetics* (22)1:59. The FGB gene encodes a 491 amino acid protein and is described in, for example, Chung, *et al.* (1983) *Ann. N. Y. Acad. Sci.* 408, 449-456.

As shown in Table 6, one polymorphism in the FGB gene found in the population screened results in a change in the amino acid sequence of the FGB protein. The FGBu4 SNP is a change from a G to an A at nucleotide residue 8059 of the reference sequence GI 182597, which results in a change from an arginine (R) to a lysine (K) at amino acid 478 of GI 11761631, the reference sequence for the FGB protein.

In one embodiment, the THBS2, ACE, or FGB polypeptides are isolated from, or otherwise substantially free of other cellular proteins. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of THBS2, ACE, or FGB polypeptides having less than about 20% (by dry weight) contaminating protein, and preferably having less than about 5% contaminating protein. It will be appreciated that functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein.

Preferred THBS2, ACE, or FGB proteins of the invention have an amino acid sequence which is at least about 60%, 70%, 80%, 85%, 90%, or 95% identical or homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, respectively. Even more preferred THBS2, ACE, or FGB proteins comprise an amino acid sequence which is at least about 95%, 96%, 97%, 98%, or 99% homologous

or identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, respectively. Such proteins can be recombinant proteins, and can be, *e.g.*, produced *in vitro* from nucleic acids comprising a specific allele of a THBS2, ACE, or FGB polymorphic region. For example, recombinant polypeptides preferred by the present invention can be encoded by a nucleic acid which comprises a sequence which is at least 85% homologous and more preferably 90% homologous and most preferably 95% homologous with a nucleotide sequence set forth in SEQ ID NOs:1, 3, or 5 and comprises an allele of a polymorphic region that differs from that set forth in SEQ ID NOs: 1, 3, or 5. Polypeptides which are encoded by a nucleic acid comprising a sequence that is at least about 98-99% homologous with the sequence of SEQ ID NOs: 1, 3, or 5 and comprises an allele of a polymorphic region that differs from that set forth in SEQ ID NOs: 1, 3, or 5 are also within the scope of the invention.

In a preferred embodiment, a THBS2, ACE, or FGB protein of the present invention is a mammalian THBS2, ACE, or FGB protein. In an even more preferred embodiment, the THBS2, ACE, or FGB protein is a human protein.

The invention also provides peptides that preferably are capable of functioning in one of either role of an agonist or antagonist of at least one biological activity of a reference ("normal") THBS2, ACE, or FGB protein of the appended sequence listing. The term "evolutionarily related to," with respect to amino acid sequences of THBS2, ACE, or FGB proteins, refers to both polypeptides having amino acid sequences found in human populations, and also to artificially produced mutational variants of human THBS2, ACE, or FGB polypeptides which are derived, for example, by combinatorial mutagenesis.

Full length proteins or fragments corresponding to one or more particular motifs and/or domains or to arbitrary sizes, for example, at least 5, 10, 25, 50, 75 and 100, amino acids in length of THBS2, ACE, or FGB protein are within the scope of the present invention.

Isolated THBS2, ACE, or FGB peptides or polypeptides can be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, such peptides and polypeptides can be chemically synthesized using techniques known in the art such as conventional

Merrifield solid phase f-Moc or t-Boc chemistry. For example, a THBS2, ACE, or FGB peptide or polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptides or polypeptides which can function as either agonists or antagonists of a wild-type (*e.g.*, "normal") THBS2, ACE, or FGB protein.

In general, peptides and polypeptides referred to herein as having an activity (*e.g.*, are "bioactive") of a THBS2, ACE, or FGB protein are defined as peptides and polypeptides which mimic or antagonize all or a portion of the biological/biochemical activities of a THBS2, ACE, or FGB protein having SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, respectively, such as the ability to bind ligands. Other biological activities of the subject THBS2, ACE, or FGB proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a peptide or polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a THBS2, ACE, or FGB protein.

Assays for determining whether a THBS2, ACE, or FGB protein or variant thereof, has one or more biological activities are well known in the art.

Other preferred proteins of the invention are those encoded by the nucleic acids set forth in the section pertaining to nucleic acids of the invention. In particular, the invention provides fusion proteins, *e.g.*, THBS2, ACE, or FGB-immunoglobulin fusion proteins. Such fusion proteins can provide, *e.g.*, enhanced stability and solubility of THBS2, ACE, or FGB proteins and may thus be useful in therapy. Fusion proteins can also be used to produce an immunogenic fragment of a THBS2, ACE, or FGB protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the THBS2, ACE, or FGB polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject THBS2, ACE, or FGB protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising THBS2, ACE, or FGB epitopes as part of the virion. It has

been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a THBS2, ACE, or FGB protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No: 0259149; and Evans *et al.* (1989) *Nature* 339:385; Huang *et al.* (1988) *J. Virol.* 62:3855; and Schlienger *et al.* (1992) *J. Virol.* 66:2).

The Multiple antigen peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a THBS2, ACE, or FGB polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett *et al.* (1988) JBC 263:1719 and Nardelli *et al.* (1992) *J. Immunol.* 148:914). Antigenic determinants of THBS2, ACE, or FGB proteins can also be expressed and presented by bacterial cells.

Fusion proteins can also facilitate the expression of proteins including the THBS2, ACE, or FGB polypeptides of the present invention. For example, THBS2, ACE, or FGB polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can be easily purified, as for example by the use of glutathione-derivatized matrices (see, for example, Current Protocols in Molecular Biology, eds. Ausubel *et al.* (N.Y.: John Wiley & Sons, 1991)) and used subsequently to yield purified THBS2, ACE, or FGB polypeptides.

The present invention further pertains to methods of producing the subject THBS2, ACE, or FGB polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the peptide to occur. Suitable media for cell culture are well known in the art. The recombinant THBS2, ACE, or FGB polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant THBS2, ACE, or FGB polypeptide

is a fusion protein containing a domain which facilitates its purification, such as GST fusion protein.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of one of the subject THBS2, ACE, or FGB polypeptides which function in a limited capacity as one of either a THBS2, ACE, or FGB agonist (mimetic) or a THBS2, ACE, or FGB antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all of the biological activities of naturally occurring forms of THBS2, ACE, or FGB proteins.

Homologs of each of the subject THBS2, ACE, or FGB proteins can be generated by mutagenesis, such as by discrete point mutation(s), and/or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the THBS2, ACE, or FGB polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to a THBS2, ACE, or FGB receptor.

The recombinant THBS2, ACE, or FGB polypeptides of the present invention also include homologs of THBS2, ACE, or FGB polypeptides which differ from the THBS2, ACE, or FGB protein having SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, respectively, such as versions of the protein which are resistant to proteolytic cleavage, as for example, due to mutations which alter ubiquitination or other enzymatic targeting associated with the protein.

THBS2, ACE, or FGB polypeptides may also be chemically modified to create THBS2, ACE, or FGB derivatives by forming covalent or aggregate conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of THBS2, ACE, or FGB proteins can be prepared by linking the chemical moieties to functional groups on amino acid side-chains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

Modification of the structure of the subject THBS2, ACE, or FGB polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (*e.g.*, ex vivo shelf life and resistance to proteolytic degradation), or post-translational modifications (*e.g.*, to alter phosphorylation pattern of protein). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the THBS2, ACE, or FGB polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition. The substitutional variant may be a substituted conserved amino acid or a substituted non-conserved amino acid.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (*i.e.*, isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional THBS2, ACE, or FGB homolog (*e.g.*, functional in the sense that the resulting polypeptide mimics or antagonizes the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively

inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

Methods

The invention further provides predictive medicine methods, which are based, at least in part, on the discovery of THBS2, ACE, or FGB polymorphic regions which are associated with specific physiological states and/or diseases or disorders, *e.g.*, vascular diseases or disorders such as CAD and MI. These methods can be used alone, or in combination with other predictive medicine methods, including the identification and analysis of known risk factors associated with vascular disease, *e.g.*, phenotypic factors such as, for example, obesity, diabetes, and/or family history.

For example, information obtained using the diagnostic assays described herein (singly or in combination with information of another genetic defect which contributes to the same disease, *e.g.*, a vascular disease or disorder) is useful for diagnosing or confirming that a subject has an allele of a polymorphic region which is associated with a particular disease or disorder, *e.g.*, a vascular disease or disorder. Moreover, the information obtained using the diagnostic assays described herein, singly or in combination with information of another genetic defect which contributes to the same disease, *e.g.*, a vascular disease or disorder, can be used to predict whether or not a subject will benefit from further diagnostic evaluation for a vascular disease or disorder. Such further diagnostic evaluation includes, but is not limited to, cardiovascular imaging, such as angiography, cardiac ultrasound, coronary angiogram, magnetic resonance imagery, nuclear imaging, CT scan, myocardial perfusion imagery, or electrocardiogram, genetic analysis, *e.g.*, identification of additional polymorphisms, *e.g.*, which contribute to the same disease, familial health history analysis, lifestyle analysis, or exercise stress tests, either alone or in combination. Furthermore, the diagnostic information obtained using the diagnostic assays described herein (singly or in combination with information of another genetic defect which contributes to the same disease, *e.g.*, a vascular disease or disorder), may be used to identify which subject will benefit from a particular clinical course of therapy useful for preventing, treating, ameliorating, or prolonging onset of the particular vascular disease or disorder in the

particular subject. Clinical courses of therapy include, but are not limited to, administration of medication, non-surgical intervention, surgical intervention or procedures, and use of surgical and non-surgical medical devices used in the treatment of vascular disease, such as, for example, stents or defibrillators.

Alternatively, the information, singly, or in combination with information of another genetic defect which contributes to the same disease, *e.g.,* a vascular disease or disorder, can be used prognostically for predicting whether a non-symptomatic subject is likely to develop a disease or condition which is associated with one or more specific alleles of THBS2, ACE, or FGB polymorphic regions in a subject. Based on the prognostic information, a health care provider can recommend a particular further diagnostic evaluation which will benefit the subject, or a particular clinical course of therapy, as described above.

In addition, knowledge of the identity of a particular THBS2, ACE, or FGB allele in a subject (the THBS2, ACE, or FGB genetic profile), singly, or in combination, allows customization of further diagnostic evaluation and/or a clinical course of therapy for a particular disease. For example, a subject's THBS2, ACE, or FGB genetic profile or the genetic profile of a disease or disorder associated with a specific allele of a THBS2, ACE, or FGB polymorphic region, *e.g.,* a vascular disease or disorder, can enable a health care provider: 1) to more efficiently and cost-effectively identify means for further diagnostic evaluation, including, but not limited to, further genetic analysis, familial health history analysis, or use of vascular imaging devices; 2) to more effectively prescribe a drug that will address the molecular basis of the disease or condition; 3) to more efficiently and cost-effectively identify an appropriate clinical course of therapy, including, but not limited to, lifestyle changes, medications, surgical or non-surgical devices, surgical or non-surgical intervention, or any combination thereof; and 4) to better determine the appropriate dosage of a particular drug or duration of a particular course of clinical therapy. For example, the expression level of THBS2, ACE, or FGB proteins, alone or in conjunction with the expression level of other genes, known to contribute to the same disease, can be measured in many subjects at various stages of the disease to generate a transcriptional or expression profile of the disease. Expression patterns of individual subjects can then be compared to the expression profile

of the disease to determine the appropriate drug, dose to administer to the subject, or course of clinical therapy.

The ability to target populations expected to show the highest clinical benefit, based on the THBS2, ACE, or FGB or disease genetic profile, can enable: 1) the repositioning of marketed drugs, surgical devices for use in treating, preventing, or ameliorating vascular diseases or disorders, or diagnostics, such as vascular imaging devices, with disappointing market results; 2) the rescue of drug candidates whose clinical development has been discontinued as a result of safety or efficacy limitations, which are subject subgroup-specific; 3) an accelerated and less costly development for drug candidates and more optimal drug labeling (*e.g.*, since the use of THBS2, ACE, or FGB as a marker is useful for optimizing effective dose); and 4) an accelerated, less costly, and more effective selection of a particular course of clinical therapy suited to a particular subject.

These and other methods are described in further detail in the following sections.

A. Prognostic and Diagnostic Assays

The present methods provide means for determining if a subject is or is not at risk of developing a disease, condition or disorder that is associated a specific THBS2, ACE, or FGB allele, *e.g.*, a vascular disease or a disease or disorder resulting therefrom.

The present invention provides methods for determining the molecular structure of a THBS2, ACE, or FGB gene, such as a human THBS2, ACE, or FGB gene, or a portion thereof. In one embodiment, determining the molecular structure of at least a portion of a THBS2, ACE, or FGB gene comprises determining the identity of an allelic variant of at least one polymorphic region of a THBS2, ACE, or FGB gene (determining the presence or absence of one or more of the allelic variants, or their complements, of SEQ ID NOs.:7, 8, 9, 10, and/or 11). A polymorphic region of a THBS2, ACE, or FGB gene can be located in an exon, an intron, at an intron/exon border, or in the 5' upstream regulatory element of the THBS2, ACE, or FGB gene.

The invention provides methods for determining whether a subject is or is not at risk of developing a disease or disorder associated with a specific allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene. Such diseases can be associated

with aberrant THBS2, ACE, or FGB activity, *e.g.*, a vascular disease or disorder such as CAD or MI.

Analysis of one or more THBS2, ACE, or FGB polymorphic regions in a subject can be useful for predicting whether a subject is or is not likely to develop a vascular disease or disorder, *e.g.*, atherosclerosis, CAD, MI, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

In preferred embodiments, the methods of the invention can be characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a specific allelic variant of one or more polymorphic regions of a THBS2, ACE, or FGB gene. Preferably, the presence of the variant allele of the THBS2, ACE, and/or FGB gene described herein are detected. The allelic differences can be: (i) a difference in the identity of at least one nucleotide or (ii) a difference in the number of nucleotides, which difference can be a single nucleotide or several nucleotides. The invention also provides methods for detecting differences in THBS2, ACE, or FGB genes such as chromosomal rearrangements, *e.g.*, chromosomal dislocation. The invention can also be used in prenatal diagnostics.

A preferred detection method is allele specific hybridization using probes overlapping the polymorphic site and having about 5, 10, 20, 25, or 30 nucleotides around the polymorphic region. In a preferred embodiment of the invention, several probes capable of hybridizing specifically to allelic variants are attached to a solid phase support, *e.g.*, a "chip". Oligonucleotides can be bound to a solid support by a variety of processes, including lithography. For example a chip can hold up to 250,000 oligonucleotides (GeneChip, Affymetrix™). Mutation detection analysis using these chips comprising oligonucleotides, also termed "DNA probe arrays" is described *e.g.*, in Cronin *et al.* (1996) Human Mutation 7:244. In one embodiment, a chip comprises all the allelic variants of at least one polymorphic region of a gene. The solid phase support is then contacted with a test nucleic acid and hybridization to the specific probes is detected. Accordingly, the identity of numerous allelic variants of one or more genes can be identified in a simple hybridization experiment. For example, the identity of the allelic variant of the nucleotide polymorphism in the 5' upstream regulatory element can be determined in a single hybridization experiment.

In other detection methods, it is necessary to first amplify at least a portion of a THBS2, ACE, or FGB gene prior to identifying the allelic variant. Amplification can be performed, *e.g.*, by PCR and/or LCR (see Wu and Wallace (1989) *Genomics* 4:560), according to methods known in the art. In one embodiment, genomic DNA of a cell is exposed to two PCR primers and amplification for a number of cycles sufficient to produce the required amount of amplified DNA. In preferred embodiments, the primers are located between 150 and 350 base pairs apart.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.*, (1988) *Bio/Technology* 6:1197), and self-sustained sequence replication (*Guatelli et al.*, (1989) *Proc. Nat. Acad. Sci.* 87:1874), and nucleic acid based sequence amplification (NABSA), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence at least a portion of a THBS2, ACE, or FGB gene and detect allelic variants, *e.g.*, mutations, by comparing the sequence of the sample sequence with the corresponding reference (control) sequence. Exemplary sequencing reactions include those based on techniques developed by Maxam and Gilbert (*Proc. Natl Acad Sci USA* (1977) 74:560) or Sanger (Sanger *et al.* (1977) *Proc. Nat. Acad. Sci* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures may be utilized when performing the subject assays (*Biotechniques* (1995) 19:448), including sequencing by mass spectrometry (see, for example, U.S. Patent Number 5,547,835 and international patent application Publication Number WO 94/16101, entitled *DNA Sequencing by Mass Spectrometry* by H. Köster; U.S. Patent Number 5,547,835 and international patent application Publication Number WO 94/21822 entitled "DNA Sequencing by Mass Spectrometry Via Exonuclease Degradation" by H. Köster), and U.S Patent Number 5,605,798 and International Patent

Application No. PCT/US96/03651 entitled *DNA Diagnostics Based on Mass Spectrometry* by H. Köster;. Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159). It will be evident to one skilled in the art that, for certain embodiments, the occurrence of only one, two or three of the nucleic acid bases need be determined in the sequencing reaction. For instance, A-track or the like, *e.g.,* where only one nucleotide is detected, can be carried out.

Yet other sequencing methods are disclosed, *e.g.,* in U.S. Patent Number 5,580,732 entitled "Method of DNA sequencing employing a mixed DNA-polymer chain probe" and U.S. Patent Number 5,571,676 entitled "Method for mismatch-directed *in vitro* DNA sequencing."

In some cases, the presence of a specific allele of a THBS2, ACE, or FGB gene in DNA from a subject can be shown by restriction enzyme analysis. For example, a specific nucleotide polymorphism can result in a nucleotide sequence comprising a restriction site which is absent from the nucleotide sequence of another allelic variant.

In a further embodiment, protection from cleavage agents (such as a nuclease, hydroxylamine or osmium tetroxide and with piperidine) can be used to detect mismatched bases in RNA/RNA DNA/DNA, or RNA/DNA heteroduplexes (Myers, *et al.* (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing a control nucleic acid, which is optionally labeled, *e.g.,* RNA or DNA, comprising a nucleotide sequence of a THBS2, ACE, or FGB allelic variant with a sample nucleic acid, *e.g.,* RNA or DNA, obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as duplexes formed based on basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine whether the control and sample nucleic acids have an identical nucleotide sequence or in which nucleotides they are different. See, for example, Cotton

et al (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control or sample nucleic acid is labeled for detection.

In another embodiment, an allelic variant can be identified by denaturing high-performance liquid chromatography (DHPLC) (Oefner and Underhill, (1995) *Am. J. Human Gen.* 57:Suppl. A266). DHPLC uses reverse-phase ion-pairing chromatography to detect the heteroduplexes that are generated during amplification of PCR fragments from individuals who are heterozygous at a particular nucleotide locus within that fragment (Oefner and Underhill (1995) *Am. J. Human Gen.* 57:Suppl. A266). In general, PCR products are produced using PCR primers flanking the DNA of interest. DHPLC analysis is carried out and the resulting chromatograms are analyzed to identify base pair alterations or deletions based on specific chromatographic profiles (see O'Donovan *et al.* (1998) *Genomics* 52:44-49).

In other embodiments, alterations in electrophoretic mobility is used to identify the type of THBS2, ACE, or FGB allelic variant. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids are denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In another preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment, the identity of an allelic variant of a polymorphic region is obtained by analyzing the movement of a nucleic acid comprising the polymorphic region in polyacrylamide gels containing a gradient of denaturant is

assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing agent gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:1275).

Examples of techniques for detecting differences of at least one nucleotide between 2 nucleic acids include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide probes may be prepared in which the known polymorphic nucleotide is placed centrally (allele-specific probes) and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al* (1989) *Proc. Natl Acad. Sci USA* 86:6230; and Wallace *et al.* (1979) *Nucl. Acids Res.* 6:3543). Such allele specific oligonucleotide hybridization techniques may be used for the simultaneous detection of several nucleotide changes in different polylmorphic regions of THBS2, ACE, or FGB. For example, oligonucleotides having nucleotide sequences of specific allelic variants are attached to a hybridizing membrane and this membrane is then hybridized with labeled sample nucleic acid. Analysis of the hybridization signal will then reveal the identity of the nucleotides of the sample nucleic acid.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the allelic variant of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238; Newton *et al.* (1989) *Nucl. Acids Res.* 17:2503). This technique is also termed "PROBE" for Probe Oligo Base Extension. In addition it may be desirable to introduce a novel restriction site in the region of the

mutation to create cleavage-based detection (Gasparini *et al* (1992) *Mol. Cell Probes* 6:1).

In another embodiment, identification of the allelic variant is carried out using an oligonucleotide ligation assay (OLA), as described, *e.g.,* in U.S. Patent Number 4,998,617 and in Landegren, U. *et al.*, (1988) *Science* 241:1077-1080. The OLA protocol uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. One of the oligonucleotides is linked to a separation marker, *e.g,*. biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate. Ligation then permits the labeled oligonucleotide to be recovered using avidin, or another biotin ligand. Nickerson, D. A. *et al.* have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, D. A. *et al.*, (1990) *Proc. Natl. Acad. Sci. (U.S.A.)* 87:8923-8927. In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Several techniques based on this OLA method have been developed and can be used to detect specific allelic variants of a polymorphic region of a THBS2, ACE, or FGB gene. For example, U.S. Patent Number 5,593,826 discloses an OLA using an oligonucleotide having 3'-amino group and a 5'-phosphorylated oligonucleotide to form a conjugate having a phosphoramidate linkage. In another variation of OLA described in Tobe *et al.* ((1996) *Nucleic Acids Res* 24: 3728), OLA combined with PCR permits typing of two alleles in a single microtiter well. By marking each of the allele-specific primers with a unique hapten, *i.e.* digoxigenin and fluorescein, each OLA reaction can be detected by using hapten specific antibodies that are labeled with different enzyme reporters, alkaline phosphatase or horseradish peroxidase. This system permits the detection of the two alleles using a high throughput format that leads to the production of two different colors.

The invention further provides methods for detecting single nucleotide polymorphisms in a THBS2, ACE, or FGB gene. Because single nucleotide polymorphisms constitute sites of variation flanked by regions of invariant sequence, their analysis requires no more than the determination of the identity of the single

nucleotide present at the site of variation and it is unnecessary to determine a complete gene sequence for each subject. Several methods have been developed to facilitate the analysis of such single nucleotide polymorphisms.

In one embodiment, the single base polymorphism can be detected by using a specialized exonuclease-resistant nucleotide, as disclosed, *e.g.*, in Mundy, C. R. (U.S. Patent Number 4,656,127). According to the method, a primer complementary to the allelic sequence immediately 3' to the polymorphic site is permitted to hybridize to a target molecule obtained from a particular animal or human. If the polymorphic site on the target molecule contains a nucleotide that is complementary to the particular exonuclease-resistant nucleotide derivative present, then that derivative will be incorporated onto the end of the hybridized primer. Such incorporation renders the primer resistant to exonuclease, and thereby permits its detection. Since the identity of the exonuclease-resistant derivative of the sample is known, a finding that the primer has become resistant to exonucleases reveals that the nucleotide present in the polymorphic site of the target molecule was complementary to that of the nucleotide derivative used in the reaction. This method has the advantage that it does not require the determination of large amounts of extraneous sequence data.

In another embodiment of the invention, a solution-based method is used for determining the identity of the nucleotide of a polymorphic site. Cohen, D. *et al.* (French Patent 2,650,840; PCT Appln. No. WO91/02087). As in the Mundy method of U.S. Patent Number 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3' to a polymorphic site. The method determines the identity of the nucleotide of that site using labeled dideoxynucleotide derivatives, which, if complementary to the nucleotide of the polymorphic site will become incorporated onto the terminus of the primer.

An alternative method, known as Genetic Bit Analysis or GBA™ is described by Goelet, P. *et al.* (PCT Appln. No. 92/15712). The method of Goelet, P. *et al.* uses mixtures of labeled terminators and a primer that is complementary to the sequence 3' to a polymorphic site. The labeled terminator that is incorporated is thus determined by, and complementary to, the nucleotide present in the polymorphic site of the target molecule being evaluated. In contrast to the method of Cohen *et al.* (French Patent

2,650,840; PCT Appln. No. WO91/02087) the method of Goelet, P. *et al.* is preferably a heterogeneous phase assay, in which the primer or the target molecule is immobilized to a solid phase.

Recently, several primer-guided nucleotide incorporation procedures for assaying polymorphic sites in DNA have been described (Komher, J. S. *et al.*, (1989) *Nucl. Acids. Res.* 17:7779-7784; Sokolov, B. P., (1990) *Nucl. Acids Res.* 18:3671; Syvanen, A. -C., *et al.*, (1990) *Genomics* 8:684-692; Kuppuswamy, M. N. *et al.*, (1991) *Proc. Natl. Acad. Sci. (U.S.A.)* 88:1143-1147; Prezant, T. R. *et al.*, (1992) *Hum. Mutat.* 1:159-164; Ugozzoli, L. *et al.*, (1992) *GATA* 9:107-112; Nyren, P. (1993) *et al.*, *Anal. Biochem.* 208:171-175). These methods differ from GBA™ in that they all rely on the incorporation of labeled deoxynucleotides to discriminate between bases at a polymorphic site. In such a format, since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A.C., *et al.*, (1993) *Amer. J. Hum. Genet.* 52:46-59).

For determining the identity of the allelic variant of a polymorphic region located in the coding region of a THBS2, ACE, or FGB gene, yet other methods than those described above can be used. For example, identification of an allelic variant which encodes a mutated THBS2, ACE, or FGB protein can be performed by using an antibody specifically recognizing the mutant protein in, *e.g.*, immunohistochemistry or immunoprecipitation. Antibodies to wild-type THBS2, ACE, or FGB or mutated forms of THBS2, ACE, or FGB proteins can be prepared according to methods known in the art.

Alternatively, one can also measure an activity of a THBS2, ACE, or FGB protein, such as binding to a THBS2, ACE, or FGB ligand. Binding assays are known in the art and involve, *e.g.*, obtaining cells from a subject, and performing binding experiments with a labeled ligand, to determine whether binding to the mutated form of the protein differs from binding to the wild-type of the protein.

Antibodies directed against reference or mutant THBS2, ACE, or FGB polypeptides or allelic variant thereof, which are discussed above, may also be used in disease diagnostics and prognostics. Such diagnostic methods, may be used to detect

abnormalities in the level of THBS2, ACE, or FGB polypeptide expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of a THBS2, ACE, or FGB polypeptide. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant THBS2, ACE, or FGB polypeptide relative to the normal THBS2, ACE, or FGB polypeptide. Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques which are well known to one of skill in the art, including but not limited to Western blot analysis. For a detailed explanation of methods for carrying out Western blot analysis, see Sambrook *et al*, 1989, supra, at Chapter 18. The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Harlow, E. and Lane, D., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), which is incorporated herein by reference in its entirety.

This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection. The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of THBS2, ACE, or FGB polypeptides. *In situ* detection may be accomplished by removing a histological specimen from a subject, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the THBS2, ACE, or FGB polypeptide, but also its distribution in the examined tissue. Using the present invention, one of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Often a solid phase support or carrier is used as a support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble

to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One means for labeling an anti-THBS2, ACE, or FGB polypeptide specific antibody is via linkage to an enzyme and use in an enzyme immunoassay (EIA) (Voller, "The Enzyme Linked Immunosorbent Assay (ELISA)", *Diagnostic Horizons* 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller, et al., (1978) *J. Clin. Pathol.* 31:507-520; Butler, (1981) *Meth. Enzymol.* 73:482-523; Maggio, (ed.) *Enzyme Immunoassay*, CRC Press, Boca Raton, FL, 1980; Ishikawa, *et al.*, (eds.) *Enzyme Immunoassay*, Kgaku Shoin, Tokyo, 1981). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type or mutant peptides through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., *Principles of*

Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in, which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

If a polymorphic region is located in an exon, either in a coding or non-coding portion of the gene, the identity of the allelic variant can be determined by determining the molecular structure of the mRNA, pre-mRNA, or cDNA. The molecular structure can be determined using any of the above described methods for determining the molecular structure of the genomic DNA, *e.g.*, see Example 1.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits, such as those described above, comprising at least one probe or primer nucleic acid described herein, which may be conveniently used, *e.g.*, to determine whether a subject is or is not at risk of developing a disease associated with a specific THBS2, ACE, or FGB allelic variant.

Sample nucleic acid to be analyzed by any of the above-described diagnostic and prognostic methods can be obtained from any cell type or tissue of a subject. For example, a subject's bodily fluid (*e.g.* blood) can be obtained by known techniques (*e.g.* venipuncture). Alternatively, nucleic acid tests can be performed on dry samples (*e.g.* hair or skin). Fetal nucleic acid samples can be obtained from maternal blood as described in International Patent Application No. WO91/07660 to Bianchi. Alternatively, amniocytes or chorionic villi may be obtained for performing prenatal testing.

Diagnostic procedures may also be performed *in situ* directly upon tissue sections (fixed and/or frozen) of subject tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such *in situ* procedures (see, for example, Nuovo, G.J., 1992, PCR *in situ* hybridization: protocols and applications, Raven Press, NY).

In addition to methods which focus primarily on the detection of one nucleic acid sequence, profiles may also be assessed in such detection schemes. Fingerprint profiles may be generated, for example, by utilizing a differential display procedure, Northern analysis and/or RT-PCR.

B. Pharmacogenomics

Knowledge of the identity of the allele of one or more THBS2, ACE, and/or FGB gene polymorphic regions in a subject (the THBS2, ACE, and/or FGB genetic profile), alone or in conjunction with information of other genetic defects associated with the same disease (the genetic profile of the particular disease) also allows selection and customization of the therapy, *e.g.*, a particular clinical course of therapy and/or further diagnostic evaluation for a particular disease to the subject's genetic profile. For example, subjects having a specific allele of a THBS2, ACE, or FGB gene, singly or in

combination, may or may not exhibit symptoms of a particular disease or be predisposed to developing symptoms of a particular disease. Further, if those subjects are symptomatic, they may or may not respond to a certain drug, *e.g.*, a specific therapeutic used in the treatment or prevention of a vascular disease or disorder, *e.g.*, CAD or MI, such as, for example, beta blocker drugs, calcium channel blocker drugs, and/or nitrate drugs, but may respond to another. Furthermore, they may or may not respond to other treatments, including, for example, use of devices for treatment of vascular disease, or surgical and/or non-surgical courses of treatment. Moreover, if a subject does or does not exhibit symptoms of a particular disease, the subject may or may not benefit from further diagnostic evaluation, including, for example, use of vascular imaging devices. Thus, generation of a THBS2, ACE, or FGB genetic profile, (*e.g.*, categorization of alterations in THBS2, ACE, or FGB genes which are associated with the development of a particular disease), from a population of subjects, who are symptomatic for a disease or condition that is caused by or contributed to by a defective and/or deficient THBS2, ACE, or FGB gene and/or protein (a THBS2, ACE, or FGB genetic population profile) and comparison of a subject's THBS2, ACE, or FGB profile to the population profile, permits the selection or design of drugs that are expected to be safe and efficacious for a particular subject or subject population (*i.e.*, a group of subjects having the same genetic alteration), as well as the selection or design of a particular clinical course of therapy or further diagnostic evaluations that are expected to be safe and efficacious for a particular subject or subject population.

For example, a THBS2, ACE, or FGB population profile can be performed by determining the THBS2, ACE, or FGB profile, *e.g.*, the identity of THBS2, ACE, or FGB alleles, in a subject population having a disease, which is associated with one or more specific alleles of THBS2, ACE, or FGB polymorphic regions. Optionally, the THBS2, ACE, or FGB population profile can further include information relating to the response of the population to a THBS2, ACE, or FGB therapeutic, using any of a variety of methods, including, monitoring: 1) the severity of symptoms associated with the THBS2, ACE, or FGB related disease; 2) THBS2, ACE, or FGB gene expression level; 3) THBS2, ACE, or FGB mRNA level; and/or 4) THBS2, ACE, or FGB protein level, and dividing or categorizing the population based on particular THBS2, ACE, or FGB

alleles. The THBS2, ACE, or FGB genetic population profile can also, optionally, indicate those particular THBS2, ACE, or FGB alleles which are present in subjects that are either responsive or non-responsive to a particular therapeutic, clinical course of therapy, or diagnostic evaluation. This information or population profile, is then useful for predicting which individuals should respond to particular drugs, particular clinical courses of therapy, or diagnostic evaluations based on their individual THBS2, ACE, or FGB genetic profile.

In a preferred embodiment, the THBS2, ACE, or FGB profile is a transcriptional or expression level profile and is comprised of determining the expression level of THBS2, ACE, or FGB proteins, alone or in conjunction with the expression level of other genes known to contribute to the same disease at various stages of the disease.

Pharmacogenomic studies can also be performed using transgenic animals. For example, one can produce transgenic mice, *e.g.,* as described herein, which contain a specific allelic variant of a THBS2, ACE, or FGB gene. These mice can be created, e.g, by replacing their wild-type THBS2, ACE, or FGB gene with an allele of the human THBS2, ACE, or FGB gene. The response of these mice to specific THBS2, ACE, or FGB particular therapeutics, clinical courses of treatment, and/or diagnostic evaluations can then be determined.

(i) Diagnostic Evaluation

In one embodiment, the polymorphisms of the present invention are used to determine the most appropriate diagnostic evaluation and to determine whether or not a subject will benefit from further diagnostic evaluation. For example, if a subject has pattern 1 or pattern 2 of the THBS2 SNPs, or the complements thereof, as described herein, that subject has a decreased risk for vascular disease. Likewise, if a subject has one copy of an A and one copy of a G at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype), or the complement thereof, that subject is at a decreased risk for vascular disease. Likewise, if a subject has two copies of a T at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, that subject is at a decreased risk for vascular disease. In addition, if a subject has one copy of a T and one copy of a C at nucleotide residue 5119 of the FGB reference

sequence GI 182597, or the complement thereof, that subject is also at a decreased risk for vascular disease. Therefore, a subject having a decreased risk for vascular disease, identified by the presence of the alleles described above, would be less likely to require or benefit from further diagnostic evaluation for a vascular disease or disorder.

Thus, in one embodiment, the invention provides methods for classifying a subject who or is or is not at risk for developing, a vascular disease or disorder as a candidate for further diagnostic evaluation for a vascular disease or disorder comprising the steps of determining the THBS2, ACE, and/or FGB genetic profile of the subject, comparing the subject's THBS2, ACE, and/or FGB genetic profile to a THBS2, ACE, and/or FGB genetic population profile, and classifying the subject based on the identified genetic profiles as a subject who is a candidate for further diagnostic evaluation for a vascular disease or disorder.

In one embodiment, the subject's THBS2, ACE, and/or FGB genetic profile is determined by identifying the nucleotide at residue 3949 and/or residue 4476 of the reference sequence GI 307505 of the THBS2 gene (polymorphism ID Nos. G5755e5 and G5755e9, respectively), the nucleotide at residue 86408 of the reference sequence GI 13027555 of the ACE gene (polymorphism ID No. G765u2), the nucleotide at residue 5119 and/or residue 8059 of the reference sequence GI 182597 of the FGB gene (polymorphism ID Nos. FGBu1 and FGBu4, respectively). Methods of further diagnostic evaluation include use of vascular imaging devices such as, for example, angiography, cardiac ultrasound, coronary angiogram, magnetic resonance imagery, nuclear imaging, CT scan, myocardial perfusion imagery, or electrocardiogram, or may include genetic analysis, familial health history analysis, lifestyle analysis, exercise stress tests, or any combination thereof.

In another embodiment, the invention provides methods for selecting an effective vascular imaging device as a diagnostic tool for a vascular disease or disorder comprising the steps of determining the THBS2, ACE, and/or FGB genetic profile of the subject; comparing the subject's THBS2, ACE, and/or FGB genetic profile to a THBS2, ACE, and/or FGB genetic population profile; and selecting an effective vascular imaging device as a diagnostic tool for a vascular disease or disorder. In a preferred embodiment, the vascular imaging device is selected from the group consisting of angiography, cardiac ultrasound, coronary angiogram, magnetic resonance imagery, nuclear imaging,

CT scan, myocardial perfusion imagery, electrocardiogram, or any combination thereof.

(ii) Clinical Course of Therapy

In another aspect, the polymorphisms of the present invention are used to determine the most appropriate clinical course of therapy for a subject who is at risk of a vascular disease or disorder, and will aid in the determination of whether the subject will benefit from such clinical course of therapy, as determined by identification of one or both of the polymorphisms of the invention.

In one aspect, the invention relates to the SNPs identified as described herein, both singly or in combination, as well as to the use of these SNPs, and others in these genes, particularly those nearby in linkage disequilibrium with these SNPs, both singly and in combination, for prediction of a particular clinical course of therapy for a subject who has, or is or is not at risk for developing, a vascular disease. In one embodiment, the invention provides a method for determining whether a subject will or will not benefit from a particular course of therapy by determining the presence of one, or both of the identities of the polymorphisms of the invention. For example, the determination of the polymorphisms of the invention, singly, or in combination, will aid in the determination of whether a subject will benefit from surgical revascularization and/or will benefit by the implantation of a stent following surgical revascularization, and will aid in the determination of the likelihood of success or failure of a particular clinical course of therapy.

For example, a subject having "pattern 1," which comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT), or the complement thereof, or "pattern 2", which comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG), or the complement thereof, is at approximately 3-fold decreased odds of vascular disease.

A subject having one copy of an A and one copy of a G at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype), or the complement thereof, is at a decreased risk for vascular disease.

A subject having two copies of a T at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is at a decreased risk for vascular disease, and a subject having one copy of a T and one copy of a C at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is also at a decreased risk for vascular disease. Also, a subject having two copies of an A at nucleotide residue 8059 of the FGB reference sequence GI 182597, or the complement thereof, is at a decreased risk for vascular disease. A subject having one copy of an A and one copy of a G at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is also at a decreased risk for vascular disease (see Example 1). Therefore, a subject with these specific alleles would be less likely to require or benefit from any clinical course of therapy.

An appropriate clinical course of therapy may include, for example, a lifestyle change, including, for example, a change in diet or environment. Other clinical courses of therapy include, but are not limited to, use of surgical procedures or medical devices. Surgical procedures used for the treatment of vascular disorders, includes, for example, surgical revascularization, such as angioplasty, *e.g.,* percutaneous transluminal coronary balloon angioplasty (PTCA), or laser angioplasty, or coronary bypass grafting (CABG). Medical devices used in the treatment or prevention of vascular diseases or disorders, include, for example, a stent, a defibrillator, a pacemaker, or any combination thereof.

C. Monitoring Effects of THBS2, ACE, or FGB Therapeutics During Clinical Trials

The present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.,* an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified, *e.g.,* by the screening assays described herein) comprising the steps of (i) obtaining a preadministration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of a THBS2, ACE, or FGB protein, mRNA or gene in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the THBS2, ACE, or FGB protein, mRNA or gene in the post-administration samples; (v) comparing the level of expression or activity of the THBS2, ACE, or FGB protein, mRNA, or gene

in the preadministration sample with those of the THBS2, ACE, or FGB protein, mRNA, or gene in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of THBS2, ACE, or FGB to higher levels than detected, *i.e.,* to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of THBS2, ACE, or FGB to lower levels than detected, *i.e.,* to decrease the effectiveness of the agent.

Cells of a subject may also be obtained before and after administration of a THBS2, ACE, or FGB therapeutic to detect the level of expression of genes other than THBS2, ACE, or FGB, to verify that the THBS2, ACE, or FGB therapeutic does not increase or decrease the expression of genes which could be deleterious. This can be done, *e.g.,* by using the method of transcriptional profiling. Thus, mRNA from cells exposed *in vivo* to a THBS2, ACE, or FGB therapeutic and mRNA from the same type of cells that were not exposed to the THBS2, ACE, or FGB therapeutic could be reverse transcribed and hybridized to a chip containing DNA from numerous genes, to thereby compare the expression of genes in cells treated and not treated with a THBS2, ACE, or FGB therapeutic. If, for example a THBS2, ACE, or FGB therapeutic turns on the expression of a proto-oncogene in a subject, use of this particular THBS2, ACE, or FGB therapeutic may be undesirable.

D. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject having or likely to develop a disorder associated with specific THBS2, ACE, or FGB alleles and/or aberrant THBS2, ACE, or FGB expression or activity, *e.g.,* vascular diseases or disorders.

i) Prophylactic Methods

In one aspect, the invention provides a method for preventing a disease or disorder associated with a specific THBS2, ACE, or FGB allele such as a vascular disease or disorder, *e.g.,* CAD or MI, and medical conditions resulting therefrom, by

administering to the subject an agent which counteracts the unfavorable biological effect of the specific THBS2, ACE, or FGB allele. Subjects at risk for such a disease can be identified by a diagnostic or prognostic assay, *e.g.*, as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms associated with specific THBS2, ACE, or FGB alleles, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the identity of the THBS2, ACE, or FGB allele in a subject, a compound that counteracts the effect of this allele is administered. The compound can be a compound modulating the activity of THBS2, ACE, or FGB, *e.g.*, a THBS2, ACE, or FGB inhibitor. The treatment can also be a specific lifestyle change, *e.g.*, a change in diet or an environmental alteration. In particular, the treatment can be undertaken prophylactically, before any other symptoms are present. Such a prophylactic treatment could thus prevent the development of aberrant vascular activity, *e.g.*, the production of atherosclerotic plaque leading to, *e.g.*, CAD or MI. The prophylactic methods are similar to therapeutic methods of the present invention and are further discussed in the following subsections.

(ii) Therapeutic Methods

The invention further provides methods of treating a subject having a disease or disorder associated with a specific allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene. Preferred diseases or disorders include vascular diseases and disorders, and disorders resulting therefrom (*e.g.*, such as, for example, atherosclerosis, CAD, MI, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism).

In one embodiment, the method comprises (a) determining the identity of an allelic variant of a one or more of a THBS2, ACE, and/or FGB; and (b) administering to the subject a compound that compensates for the effect of the specific allelic variant(s). The polymorphic region can be localized at any location of the gene, *e.g.*, in a regulatory element (*e.g.*, in a 5' upstream regulatory element), in an exon, (*e.g.*, coding region of an exon), in the 3' UTR, in an intron, or at an exon/intron border. Thus, depending on the site of the polymorphism in the THBS2, ACE, or FGB gene, a subject having a specific variant of the polymorphic region which is associated with a specific disease or

condition, can be treated with compounds which specifically compensate for the effect of the allelic variant.

In a preferred embodiment, the identity of one or more of the following nucleotides of a THBS2, ACE, or FGB gene of a subject is determined: the nucleotide at residue 3949 and/or residue 4476 of the reference sequence GI 307505 of the THBS2 gene (polymorphism ID Nos. G5755e5 and G5755e9, respectively), the nucleotide at residue 86408 of the reference sequence GI 13027555 of the ACE gene (polymorphism ID No. G765u2), the nucleotide at residue 5119 and/or residue 8059 of the reference sequence GI 182597 of the FGB gene (polymorphism ID Nos. FGBu1 and FGBu4, respectively). In a preferred embodiment, the identities of one or more nucleotides is determined.

For example, a subject having "pattern 1," which comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT), or the complement thereof, or "pattern 2", which comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG), or the complement thereof, is at approximately 3-fold decreased odds of vascular disease.

A subject having one copy of an A and one copy of a G at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype), or the complement thereof, is at a decreased risk for vascular disease.

A subject having two copies of a T at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is at a decreased risk for vascular disease, and a subject having one copy of a T and one copy of a C at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is also at a decreased risk for vascular disease. Also, a subject having two copies of an A at nucleotide residue 8059 of the FGB reference sequence GI 182597, or the complement thereof, is at a decreased risk for vascular disease. A subject having one copy of an A and one copy of a G at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, is also at a decreased risk for vascular disease.

Generally, the allelic variant can be a mutant allele, *i.e.*, an allele which when present in one, or two copies, in a subject results in a change in the phenotype of the subject. A mutation can be a substitution, deletion, and/or addition of at least one nucleotide relative to the wild-type allele (*i.e.*, the reference sequence). Depending on where the mutation is located in the THBS2, ACE, or FGB gene, the subject can be treated to specifically compensate for the mutation. For example, if the mutation is present in the coding region of the gene and results in a more active THBS2, ACE, or FGB protein, the subject can be treated, *e.g.*, by administration to the subject of a medication or course of clinical treatment which treat, prevents, or ameliorates a vascular disease or disorder. Normal THBS2, ACE, or FGB protein can also be used to counteract or compensate for the endogenous mutated form of the THBS2, ACE, or FGB protein. Normal THBS2, ACE, or FGB protein can be directly delivered to the subject or indirectly by gene therapy wherein some cells in the subject are transformed or transfected with an expression construct encoding wild-type THBS2, ACE, or FGB protein. Nucleic acids encoding reference human THBS2, ACE, or FGB protein are set forth in SEQ ID NOs.:1, 3, and 5, respectively (GI Accession Nos. 307505, 13027555, and 182597, respectively).

Yet in another embodiment, the invention provides methods for treating a subject having a mutated THBS2, ACE, or FGB gene, in which the mutation is located in a regulatory region of the gene. Such a regulatory region can be localized in the 5' upstream regulatory element of the gene, in the 5' or 3' untranslated region of an exon, or in an intron. A mutation in a regulatory region can result in increased production of THBS2, ACE, or FGB protein, decreased production of THBS2, ACE, or FGB protein, or production of THBS2, ACE, or FGB having an aberrant tissue distribution. The effect of a mutation in a regulatory region upon the THBS2, ACE, or FGB protein can be determined, *e.g.*, by measuring the THBS2, ACE, or FGB protein level or mRNA level in cells having a THBS2, ACE, or FGB gene having this mutation and which, normally (*i.e.*, in the absence of the mutation) produce THBS2, ACE, or FGB protein. The effect of a mutation can also be determined *in vitro*. For example, if the mutation is in the 5' upstream regulatory element, a reporter construct can be constructed which comprises the mutated 5' upstream regulatory element linked to a reporter gene, the

construct transfected into cells, and comparison of the level of expression of the reporter gene under the control of the mutated 5' upstream regulatory element and under the control of a wild-type 5' upstream regulatory element. Such experiments can also be carried out in mice transgenic for the mutated 5' upstream regulatory element. If the mutation is located in an intron, the effect of the mutation can be determined, *e.g.*, by producing transgenic animals in which the mutated THBS2, ACE, or FGB gene has been introduced and in which the wild-type gene may have been knocked out. Comparison of the level of expression of THBS2, ACE, or FGB in the mice transgenic for the mutant human THBS2, ACE, or FGB gene with mice transgenic for a wild-type human THBS2, ACE, or FGB gene will reveal whether the mutation results in increased, or decreased synthesis of the THBS2, ACE, or FGB protein and/or aberrant tissue distribution of THBS2, ACE, or FGB protein. Such analysis could also be performed in cultured cells, in which the human mutant THBS2, ACE, or FGB gene is introduced and, *e.g.*, replaces the endogenous wild-type THBS2, ACE, or FGB gene in the cell. Thus, depending on the effect of the mutation in a regulatory region of a THBS2, ACE, or FGB gene, a specific treatment can be administered to a subject having such a mutation. Accordingly, if the mutation results in increased THBS2, ACE, or FGB protein levels, the subject can be treated by administration of a compound which reduces THBS2, ACE, or FGB protein production, *e.g.*, by reducing THBS2, ACE, or FGB gene expression or a compound which inhibits or reduces the activity of THBS2, ACE, or FGB.

A correlation between drug responses and specific alleles of THBS2, ACE, or FGB can be shown, for example, by clinical studies wherein the response to specific drugs of subjects having different allelic variants of a polymorphic region of a THBS2, ACE, or FGB gene is compared. Such studies can also be performed using animal models, such as mice having various alleles of human THBS2, ACE, or FGB genes and in which, *e.g.*, the endogenous THBS2, ACE, or FGB has been inactivated such as by a knock-out mutation. Test drugs are then administered to the mice having different human THBS2, ACE, or FGB alleles and the response of the different mice to a specific compound is compared. Accordingly, the invention provides assays for identifying the drug which will be best suited for treating a specific disease or condition in a subject.

For example, it will be possible to select drugs which will be devoid of toxicity, or have the lowest level of toxicity possible for treating a subject having a disease or condition.

Other Uses For the Nucleic Acid Molecules of the Invention

The identification of different alleles of THBS2, ACE, or FGB can also be useful for identifying an individual among other individuals from the same species. For example, DNA sequences can be used as a fingerprint for detection of different individuals within the same species (Thompson, J. S. and Thompson, eds., Genetics in Medicine, WB Saunders Co., Philadelphia, PA (1991)). This is useful, for example, in forensic studies and paternity testing, as described below.

A. Forensics

Determination of which specific allele occupies a set of one or more polymorphic sites in an individual identifies a set of polymorphic forms that distinguish the individual from others in the population. *See generally* National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard *et al.*, National Academy Press, DC, 1996). The more polymorphic sites that are analyzed, the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, the polymorphisms of the invention can be used in conjunction with known polymorphisms in distal genes. Preferred polymorphisms for use in forensics are biallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of polymorphic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers is the same in the sample as in the suspect, one can conclude that the DNA from the suspect is

consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (*e.g.*, by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

p(ID) is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. For example, in biallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y, the probability of each genotype in a diploid organism is (see WO 95/12607):

Homozygote: $p(AA)= x^2$

Homozygote: $p(BB)= y^2 = (1-x)^2$

Single Heterozygote: $p(AB)= p(BA)= xy = x(1-x)$

Both Heterozygotes: $p(AB+BA)= 2xy = 2x(1-x)$

The probability of identity at one locus (*i.e.*, the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation: $p(ID) = (x^2)$.

These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity p(ID) for a 3-allele system where the alleles have the frequencies in the population of x, y, and z, respectively, is equal to the sum of the squares of the genotype frequencies: $P(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$.

In a locus of n alleles, the appropriate binomial expansion is used to calculate p(ID) and p(exc).

The cumulative probability of identity (cum p(ID)) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus: cum $p(ID) = p(ID1)p(ID2)p(ID3)...p(IDn)$.

The cumulative probability of non-identity for n loci (*i.e.*, the probability that two random individuals will be difference at 1 or more loci) is given by the equation: cum p(nonID) = 1-cum p(ID).

If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (*e.g.*, one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known, and thus, it is possible to trace the mother's contribution to the child's genotype. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent to that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and in the child.

If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that that putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of a coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607): p(exc) = xy(1-xy), where x and y are the population frequencies of alleles A and B of a biallelic polymorphic site.

(At a triallelic site p(exc) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)), where x, y, and z and the respective populations frequencies of alleles A, B, and C).

The probability of non-exclusion is: p(non-exc) = 1-p(exc).

The cumulative probability of non-exclusion (representing the values obtained when n loci are is used) is thus:

Cum p(non-exc) = p(non-exc1)p(non-exc2)p(non-exc3)...p(non-excn).

The cumulative probability of the exclusion for n loci (representing the probability that a random male will be excluded: cum p(exc) = 1 – cum p(non-exc).

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his or her father.

C. Kits

As set forth herein, the invention provides methods, *e.g.*, diagnostic and therapeutic methods, *e.g.*, for determining the type of allelic variant of a polymorphic region present in a THBS2, ACE, or FGB gene, such as a human THBS2, ACE, or FGB gene. In preferred embodiments, the methods use probes or primers comprising nucleotide sequences which are complementary polymorphic region of a THBS2, ACE, or FGB gene (SEQ ID NOs:5, 6, 7, 8, 9, 10, and 11). Accordingly, the invention provides kits for performing these methods.

In a preferred embodiment, the invention provides a kit for determining whether a subject is or is not at risk of developing a disease or condition associated with a specific allelic variant of a THBS2, ACE, or FGB polymorphic region. In an even more preferred embodiment, the disease or disorder is characterized by an abnormal THBS2, ACE, or FGB activity. In an even more preferred embodiment, the invention provides a kit for determining whether a subject is or is not at risk of developing a vascular disease, *e.g.*, atherosclerosis, CAD, MI, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

A preferred kit provides reagents for determining whether a subject is or is not likely to develop a vascular disease, *e.g.*, CAD or MI.

Preferred kits comprise at least one probe or primer which is capable of specifically hybridizing under stringent conditions to a THBS2, ACE, or FGB reference sequence or polymorphic region and instructions for use. The kits preferably comprise at least one of the above described nucleic acids. Preferred kits for amplifying at least a portion of a THBS2, ACE, or FGB gene, comprise at least one primer pair which is capable of hybridizing to an allelic variant sequence of a THBS2, ACE, or FGB gene. The kits of the invention can also comprise one or more control nucleic acids or reference nucleic acids. For example, a kit can comprise primers for amplifying a

polymorphic region of a THBS2, ACE, or FGB gene and a control DNA corresponding to such an amplified DNA and having the nucleotide sequence of a specific allelic variant. Thus, direct comparison can be performed between the DNA amplified from a subject and the DNA having the nucleotide sequence of a specific allelic variant. In one embodiment, the control nucleic acid comprises at least a portion of a THBS2, ACE, or FGB gene of an individual who does not have a vascular disease, or a disease or disorder associated with an aberrant THBS2, ACE, or FGB activity. In another embodiment, the control nucleic acid comprises at least a portion of a THBS2, ACE, or FGB gene of an individual who does have a vascular disease, or a disease or disorder associated with an aberrant THBS2, ACE, or FGB activity. In yet another embodiment, the control nucleic acid comprises a reference sequence of a THBS2, ACE, or FGB gene.

Yet other kits of the invention comprise at least one reagent necessary to perform the assay. For example, the kit can comprise an enzyme. Alternatively the kit can comprise a buffer or any other necessary reagent.

D. Electronic Apparatus Readable Media and Arrays

Electronic apparatus readable media comprising a polymorphism of the present invention is also provided. As used herein, "electronic apparatus readable media" and "computer readable media," which are used interchangeably herein, refer to any suitable medium for storing, holding or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon a polymorphism of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet;

electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the polymorphisms of the present invention.

A variety of software programs and formats can be used to store the polymorphism information of the present invention on the electronic apparatus readable medium. For example, the polymorphic sequence can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of data processor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the markers of the present invention.

By providing the polymorphisms of the invention in readable form, one can routinely access the polymorphism information for a variety of purposes. For example, one skilled in the art can use the sequences of the polymorphisms of the present invention in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

The present invention therefore provides a medium for holding instructions for performing a method for determining whether or not a subject has a vascular disease or a pre-disposition to a vascular disease, wherein the method comprises the steps of determining the presence or absence of a polymorphism and based on the presence or absence of the polymorphism, determining whether the subject has a vascular disease or a pre-disposition to a vascular disease and/or recommending a particular clinical course of therapy or diagnostic evaluation for the vascular disease or pre-vascular disease condition.

The present invention further provides in an electronic system comprising a processor and/or in a network, a method for determining whether or not a subject has a vascular disease or a pre-disposition to vascular disease associated with a polymorphism as described herein wherein the method comprises the steps of determining the presence or absence of the polymorphism, and based on the presence or absence of the polymorphism, determining whether the subject has a vascular disease or a pre-disposition to a vascular disease, and/or recommending a particular treatment for the vascular disease or pre-vascular disease condition. In one embodiment, the processor implements the functionality of obtaining information from the subject indicative of the presence or absence of the polymorphic region. In another embodiment, the processor further implements the functionality of receiving phenotypic information associated with the subject. In yet another embodiment, the processor further implements the functionality of acquiring from a network phenotypic information associated with the subject. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

The present invention also provides in a network, a method for determining whether or not a subject has vascular disease or a pre-disposition to vascular disease associated with a polymorphism, said method comprising the steps of receiving information associated with the polymorphism, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to the polymorphism and/or vascular disease, and based on one or more of the phenotypic information, the polymorphism, and the acquired information, determining whether or not the subject has a vascular disease or a pre-disposition to a vascular disease. The method may further comprise the step of recommending a particular treatment for the vascular disease or pre-vascular disease condition.

The present invention also provides a method for determining whether or not a subject has a vascular disease or a pre-disposition to a vascular disease, said method comprising the steps of receiving information associated with the polymorphism, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to the polymorphism and/or vascular disease, and based

on one or more of the phenotypic information, the polymorphism, and the acquired information, determining whether the subject has vascular disease or a pre-disposition to vascular disease. The method may further comprise the step of recommending a particular treatment for the vascular disease or pre-vascular disease condition.

E. Personalized Health Assessment

Methods and systems of assessing personal health and risk for disease, *e.g.*, vascular disease, in a subject, using the polymorphisms and associations of the instant invention are also provided. The methods provide personalized health care knowledge to individuals as well as to their health care providers, as well as to health care companies. , It will be appreciated that the term "health care providers" is not limited to physicians but can be any source of health care. The methods and systems provide personalized information including a personal health assessment report that can include a personalized molecular profile, *e.g.*, a THBS2, ACE, and/or FGB genetic profile, a health profile, or both. Overall, the methods and systems as described herein provide personalized information for individuals and patient management tools for healthcare providers and/or subjects using a variety of communications networks such as, for example, the Internet. U.S. Patent Application Serial No. 60/266,082, filed February 1, 2001, entitled "Methods and Systems for Personalized Health Assessment," further describes personalized health assessment methods, systems, and apparatus, and is expressly incorporated herein by reference.

In one aspect, the invention provides an Internet-based method for assessing a subject's risk for vascular disease, *e.g.*, CAD or MI. In one embodiment, the method comprises obtaining information from the subject regarding the polymorphic region of an F7 gene, through *e.g.*, obtaining a biological sample from a subject, analyzing the biological sample to determine the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB, and providing results of the analysis to the subject via the Internet, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease. In another embodiment, the method comprises analyzing data from a biological sample from a subject relating to the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB and

providing results of the analysis to the subject via the Internet, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates an a decreased risk for vascular disease.

It will be appreciated that the phrase "wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease" includes a subject having "pattern 1," which comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT), or the complement thereof, or "pattern 2", which comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG), or the complement thereof, which indicates that the subject is at approximately 3-fold decreased odds of having or developing a vascular disease. This phrase also includes a subject having one copy of an A and one copy of a G at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype), or the complement thereof, which indicates that the subject is at a decreased risk for having or developing a vascular disease. This phrase also includes a subject having two copies of a T at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, which indicates that the subject is at a decreased risk for having or developing a vascular disease, and a subject having one copy of a T and one copy of a C at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, which indicates that the subject is also at a decreased risk for having or developing a vascular disease. Also, a subject having two copies of an A at nucleotide residue 8059 of the FGB reference sequence GI 182597, or the complement thereof, indicates that the subject is at a decreased risk for having or developing a vascular disease. A subject having one copy of an A and one copy of a G at nucleotide residue 5119 of the FGB reference sequence GI 182597, or the complement thereof, indicates that the subject is also at a decreased risk for having or developing a vascular disease (see Example 1).

The terms "Internet" and/or "communications network" as used herein refer to any suitable communication link, which permits electronic communications. It should be understood that these terms are not limited to "the Internet" or any other particular system or type of communication link. That is, the terms "Internet" and/or "communications network" refer to any suitable communication system, including extra-

computer system and intra-computer system communications. Examples of such communication systems include internal busses, local area networks, wide area networks, point-to-point shared and dedicated communications, infra-red links, microwave links, telephone links, CATV links, satellite and radio links, and fiber-optic links. The terms "Internet" and/or "communications network" can also refer to any suitable communications system for sending messages between remote locations, directly or via a third party communication provider such as AT&T. In this instance, messages can be communicated via telephone or facsimile or computer synthesized voice telephone messages with or without voice or tone recognition, or any other suitable communications technique.

In another aspect, the methods of the invention also provide methods of assessing a subject's risk for vascular disease, *e.g.*, CAD or MI. In one embodiment, the method comprises obtaining information from the subject regarding the polymorphic region of an F7 gene, through *e.g.*, obtaining a biological sample from the individual, analyzing the sample to obtain the subject's THBS2, ACE, and/or FGB genetic profile, representing the THBS2, ACE, and/or FGB genetic profile information as digital genetic profile data, electronically processing the THBS2, ACE, and/or FGB digital genetic profile data to generate a risk assessment report for vascular disease, and displaying the risk assessment report on an output device, where the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease. In another embodiment, the method comprises analyzing a subject's THBS2, ACE, and/or FGB genetic profile, representing the THBS2, ACE, and/or FGB genetic profile information as digital genetic profile data, electronically processing the THBS2, ACE, and/or FGB digital genetic profile data to generate a risk assessment report for vascular disease, and displaying the risk assessment report on an output device, where the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease, *e.g.*, CAD or MI. Additional health information may be provided and can be utilized to generate the risk assessment report. Such information includes, but is not limited to, information regarding one or more of age, sex, ethnic origin, diet, sibling health, parental health, clinical symptoms, personal health history, blood test data, weight, and alcohol use, drug use, nicotine use, and blood pressure.

The THBS2, ACE, and/or FGB digital genetic profile data may be transmitted via a communications network, *e.g.*, the Internet, to a medical information system for processing.

In yet another aspect the invention provides a medical information system for assessing a subject's risk for vascular disease comprising a means for obtaining information from the subject regarding the polymorphic region of an F7 gene, through *e.g.*, obtaining a biological sample from the individual to obtain a THBS2, ACE, and/or FGB genetic profile, a means for representing the THBS2, ACE, and/or FGB genetic profile as digital molecular data, a means for electronically processing the THBS2, ACE, and/or FGB digital genetic profile to generate a risk assessment report for vascular disease, and a means for displaying the risk assessment report on an output device, where the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

In another aspect, the invention provides a computerized method of providing medical advice to a subject comprising obtaining information from the subject regarding the polymorphic region of an F7 gene, through *e.g.*, obtaining a biological sample from the subject, analyzing the subject's biological sample to determine the subject's THBS2, ACE, and/or FGB genetic profile, and, based on the subject's THBS2, ACE, and/or FGB genetic profile, determining the subject's risk for vascular disease. Medical advice may be then provided electronically to the subject, based on the subject's risk for vascular disease. The medical advice may comprise, for example, recommending one or more of the group consisting of: further diagnostic evaluation, use of medical or surgical devices, administration of medication, or lifestyle change. Additional health information may also be obtained from the subject and may also be used to provide the medical advice.

In another aspect, the invention includes a method for self-assessing risk for a vascular disease. The method comprises providing information from the subject regarding the polymorphic region of an F7 gene, through *e.g.*, providing a biological sample for genetic analysis, and accessing an electronic output device displaying results of the genetic analysis, thereby self-assessing risk for a vascular disease, where the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

In another aspect, the invention provides a method of self-assessing risk for vascular disease comprising providing information from the subject regarding the polymorphic region of an F7 gene, through *e.g.*, providing a biological sample, accessing THBS2, ACE, and/or FGB digital genetic profile data obtained from the biological sample, the THBS2, ACE, and/or FGB digital genetic profile data being displayed via an output device, where the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

An output device may be, for example, a CRT, printer, or website. An electronic output device may be accessed via the Internet.

The biological sample may be obtained from the individual at a laboratory company. In one embodiment, the laboratory company processes the biological sample to obtain THBS2, ACE, and/or FGB genetic profile data, represents at least some of the THBS2, ACE, and/or FGB genetic profile data as digital genetic profile data, and transmits the THBS2, ACE, and/or FGB digital genetic profile data via a communications network to a medical information system for processing. The biological sample may also be obtained from the subject at a draw station. A draw station processes the biological sample to obtain THBS2, ACE, and/or FGB genetic profile data and transfers the data to a laboratory company. The laboratory company then represents at least some of the THBS2, ACE, and/or FGB genetic profile data as digital genetic profile data, and transmits the THBS2, ACE, and/or FGB digital genetic profile data via a communications network to a medical information system for processing.

In another aspect, the invention provides a method for a health care provider to generate a personal health assessment report for an individual. The method comprises counseling the individual to provide a biological sample and authorizing a draw station to take a biological sample from the individual and transmit molecular information from the sample to a laboratory company, where the molecular information comprises the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB. The health care provider then requests the laboratory company to provide digital molecular data corresponding to the molecular information to a medical information system to electronically process the digital molecular data and digital health data obtained from the

individual to generate a health assessment report, receives the health assessment report from the medical information system, and provides the health assessment report to the individual.

In still another aspect, the invention provides a method of assessing the health of an individual. The method comprises obtaining health information from the individual using an input device (*e.g.,* a keyboard, touch screen, hand-held device, telephone, wireless input device, or interactive page on a website), representing at least some of the health information as digital health data, obtaining information from the subject regarding the polymorphic region of an F7 gene, through *e.g.,* obtaining a biological sample from the individual, and processing the biological sample to obtain molecular information, where the molecular information comprises the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB. At least some of the molecular information and health data is then presented as digital molecular data and electronically processed to generate a health assessment report. The health assessment report is then displayed on an output device. The health assessment report can comprise a digital health profile of the individual. The molecular data can comprise protein sequence data, and the molecular profile can comprise a proteomic profile. The molecular data can also comprise information regarding one or more of the absence, presence, or level, of one or more specific proteins, polypeptides, chemicals, cells, organisms, or compounds in the individual's biological sample. The molecular data may also comprise, *e.g.,* nucleic acid sequence data, and the molecular profile may comprise, *e.g.,* a genetic profile.

In yet another embodiment, the method of assessing the health of an individual further comprises obtaining a second biological sample or a second health information at a time after obtaining the initial biological sample or initial health information, processing the second biological sample to obtain second molecular information, processing the second health information, representing at least some of the second molecular information as digital second molecular data and second health information as digital health information, and processing the molecular data and second molecular data and health information and second health information to generate a health assessment report. In one embodiment, the health assessment report provides information about the individual's predisposition for vascular disease, *e.g.,* CAD or MI, and options for risk

reduction.

Options for risk reduction comprise, for example, one or more of diet, exercise, one or more vitamins, one or more drugs, cessation of nicotine use, and cessation of alcohol use. wherein the health assessment report provides information about treatment options for a particular disorder. Treatment options comprise, for example, one or more of diet, one or more drugs, physical therapy, and surgery. In one embodiment, the health assessment report provides information about the efficacy of a particular treatment regimen and options for therapy adjustment.

In another embodiment, electronically processing the digital molecular data and digital health data to generate a health assessment report comprises using the digital molecular data and/or digital health data as inputs for an algorithm or a rule-based system that determines whether the individual is at risk for a specific disorder, *e.g.*, a vascular disorder, such as CAD or MI. Electronically processing the digital molecular data and digital health data may also comprise using the digital molecular data and digital health data as inputs for an algorithm or a rule-based system based on one or more databases comprising stored digital molecular data and/or digital health data relating to one or more disorders, *e.g.*, vascular disorders, such as CAD or MI.

In another embodiment, processing the digital molecular data and digital health data comprises using the digital molecular data and digital health data as inputs for an algorithm or a rule-based system based on one or more databases comprising: (i) stored digital molecular data and/or digital health data from a plurality of healthy individuals, and (ii) stored digital molecular data and/or digital health data from one or more pluralities of unhealthy individuals, each plurality of individuals having a specific disorder. At least one of the databases can be a public database. In one embodiment, the digital health data and digital molecular data are transmitted via, *e.g.*, a communications network, *e.g.*, the Internet, to a medical information system for processing.

A database of stored molecular data and health data, *e.g.*, stored digital molecular data and/or digital health data, from a plurality of individuals, is further provided. A database of stored digital molecular data and/or digital health data from a plurality of healthy individuals, and stored digital molecular data and/or digital health data from one or more pluralities of unhealthy individuals, each plurality of individuals having a

specific disorder, *e.g.*, a vascular disorder, is also provided.

The new methods and systems of the invention provide healthcare providers with access to ever-growing relational databases that include both molecular data and health data that is linked to specific disorders, *e.g.*, vascular disorders. In addition public medical knowledge is screened and abstracted to provide concise, accurate information that is added to the database on an ongoing basis. In addition, new relationships between particular SNPs, *e.g.*, SNPs associated with vascular disease, or genetic mutations and specific discords are added as they are discovered.

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention. The contents of all references, issued patents and published patent applications cited throughout this application, as well as the Figures, Tables, and database references, including GenBank Accession Numbers, are incorporated herein by reference. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis *et al.* U.S. Patent Number 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu *et al.* eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986);

Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

EXAMPLES

Example 1: Detection of polymorphic regions in the human THBS2, ACE, and FGB genes

This example describes the detection of polymorphic regions in the human THBS2, ACE, and FGB genes through use of denaturing high performance liquid chromatography (DHPLC), variant detector arrays, polymerase chain reaction (PCR), and direct sequencing.

Cell lines derived from an ethnically diverse population were obtained and used for single nucleotide polymorphism (SNP) discovery by methods described in Cargill, *et al.* (1999) *Nature Genetics* 22:231-238.

Genomic sequence representing the coding and partial regulatory regions of genes were amplified by polymerase chain reaction and screened via two independent methods: denaturing high performance liquid chromatography (DHPLC) or variant detector arrays (Affymetrix™).

DHPLC uses reverse-phase ion-pairing chromatography to detect the heteroduplexes that are generated during amplification of PCR fragments from individuals who are heterozygous at a particular nucleotide locus within that fragment (Oefner and Underhill (1995) *Am. J. Human Gen.* 57:Suppl. A266).

Generally, the analysis was carried out as described in O'Donovan *et al.* ((1998) Genomics 52:44-49). PCR products having product sizes ranging from about 150-400 bp were generated using the primers and PCR conditions described in Example 2. Two PCR reactions were pooled together for DHPLC analysis (4 ul of each reaction for a total of 8 ul per sample). DHPLC was performed on a DHPLC system purchased from Transgenomic, Inc. The gradient was created by mixing buffers A (0.1M TEAA) and B (0.1M TEAA, 25% Acetontitrile). WAVEmaker™ software was utilized to predict a melting temperature and calculate a buffer gradient for mutation analysis of a given DNA sequence. The resulting chromatograms were analyzed to identify base pair

alterations or deletions based on specific chromatographic profiles.

Detection of polymorphic regions in the human THBS2, ACE, and FGB genes by SSCP

Genomic DNA from the cell lines derived from an ethnically diverse population as described in Cargill, *et al.* (1999) *Nature Genetics* 22:231-238, was subjected to PCR in 25 µl reactions (1X PCR Amplitaq polymerase buffer, 0.1 mM dNTPs, 0.8 µM 5' primer, 0.8 µM 3' primer, 0.75 units of Amplitaq polymerase, 50 ng genomic DNA) using each of the above described pairs of primers under the following cycle conditions: 94°C for 2 min, 35 x [94°C for 40 sec, 57°C for 30 sec, 72°C for 1 min], 72°C 5 min, 4°C hold.

The amplified genomic DNA fragments were then analyzed by SSCP (Orita *et al.* (1989) *PNAS USA* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). From each 25 µl PCR reaction, 3 µl was taken and added to 7 µl of loading buffer. The mixture was heated to 94°C for 5 min and then immediately cooled in a slurry of ice-water. 3-4 µl were then loaded on a 10% polyacrylamide gel either with 10% glycerol or without 10% glycerol, and then subjected to electrophoresis either overnight at 4 Watts at room temperature, overnight at 4 Watts at 4°C (for amplifying a 5' upstream regulatory element), or for 5 hours at 20 Watts at 4°C. The secondary structure of single-stranded nucleic acids varies according to sequence, thus allowing the detection of small differences in nucleic acid sequence between similar nucleic acids. At the end of the electrophoretic period, the DNA was analyzed by gently overlaying a mixture of dyes onto the gel (1x the manufacturer's recommended concentration of SYBR Green I™ and SYBR Green II™ in 0.5 X TBE buffer (Molecular Probes™)) for 5 min, followed by rinsing in distilled water and detection in a Fluoroimager 575™ (Molecular Dynamics™).

Identification of polymorphic regions in the human THBS2, ACE, or FGB gene by direct sequencing of PCR products

To determine the sequences of the polymorphisms identified, the regions containing the polymorphisms were reamplified using flanking primers. The genomic DNA was subjected to PCR in 50 µl reactions (1x PCR Amplitaq polymerase buffer, 0.1 mM dNTPs, 0.8 µM 5' primer, 0.8 µM 3' primer, 0.75 units of Amplitaq polymerase, 50 ng genomic DNA) using each of the pairs of primers under the following cycle conditions: 94°C for 2 min, 35 x [94°C for 40 sec, 57°C for 30 sec, 72°C for 1 min], 72°C 5 min, 4°C hold. The newly amplified products were then purified using the Qiagen Qiaquick PCR purification kit according to the manufacturer's protocol, and subjected to sequencing using the aforementioned primers which were utilized for amplification.

Case-Control population

Several SNPs in each of the THBS2, ACE, and FGB genes were identified. Further analysis of the THBS2, ACE, and FGB SNPs included genotyping of the SNPs in large patient populations to assess their association with CAD and MI. A total of 352 U.S. Caucasian subjects with premature coronary artery disease were identified in 15 participating medical centers, fulfilling the criteria of either myocardial infarction, surgical or percutaneous revascularization, or a significant coronary artery lesion (*e.g.*, at least a 70% stenosis in a major epicardial artery) diagnosed before age 45 in men or age 50 in women and having a living sibling who met the same criteria. The sibling with the earliest onset in a Caucasian subset of these families was compared with a random sample of 418 Caucasian controls without known coronary disease. Controls representing a general, unselected population were identified through random-digit dialing in the Atlanta, Georgia area. Subjects ranging in age from age 20 to age 70 were invited to participate in the study. The subjects answered a health questionnaire, had anthropometric measures taken, and blood drawn for measurement of serum markers and extraction of DNA.

Statistical Analysis

All analyses were done using the SAS statistical package (Version 8.0, SAS Institute Inc., Cary, N.C.). Differences between cases and controls were assessed with a chi-square statistic for categorical covariates and the Wilcoxon statistic for continuous covariates. Association between each SNP and two outcomes, CAD and MI, was measured by comparing genotype frequencies between controls and all CAD cases and the subset of cases with MI. Significance was determined using a continuity-adjusted chi-square or Fisher's exact test for each genotype compared to the homozygotes wild-type for that locus. Odds ratios were calculated and presented with 95% confidence intervals.

Genotype groups were pooled for subsequence analysis of the top loci. Pooling allows the best model for each locus (dominant, codominant, or recessive) to be tested. Models were chosen based on significant differences between genotypes within a locus. A recessive model was chosen when the homozygous variant differed significantly from both the heterozygous and homozygous wildtype, and the latter two did not differ from each other. A codominant model was chosen when homozygous variant genotypes differed from both heterozygous and homozygous wild-type, and the latter two differed significantly from each other. A dominant model was chosen when no significant difference was observed between heterozygous and homozygous variant genotypes.

Multivariate logistic regression was used to adjust for sex, presence of hypertension, diabetes, and body mass index using the LOGISTC procedure in SAS. Height and weight, measured at the time of enrollment, were used to calculate body mass index for each subject. Presence of hypertension and non-insulin-dependent diabetes was measures by self-report (controls) and medical record confirmation (cases).

Results: Identified SNPs and Associations with Vascular Disease

THBS2

Two SNPs in the THBS2 gene were identified and found to be associated with vascular disease, *e.g.,* CAD and MI. The first THBS2 SNP, referred to herein as G5755e5, is a change from the thymidine (T) to a guanine (G) in the THBS2 gene at

residue 3949 of the reference sequence GI 307505. The second THBS2 SNP, referred to herein as G5755e9, is a change from a thymidine (T) to a cytidine (C) in the THBS2 gene at residue 4476 of the reference sequence GI 307505. These SNPs are located within the 3' untranslated region of the THBS2 gene. Therefore, they do not result in a change in the amino acid sequence of the THBS2 protein (see Table 1, below).

Table 1.

1	2	3	4	5	6	7	8	9	10
Gene	**PolyID**	**variant freq.**	**Type of var.**	**Geno-types**	**Ref.**	**Var.**	**Genbank Accession No./nt position**	**Flanking sequence**	**SEQ ID NO:**
THBS2	G5755e5	.29	3'	GG GT TT	T	G	GI: 307505/ nt 3949	AATGGAA CgCAGAG ATG	7
THBS2	G5755e9	.13	3' utr	CC CT TT	T	C	GI: 307505/ nt 4476	TGCAAAT GGGTGTG AcGCGGT TCCAGAT GTG	8

The variant allele, G, of the THBS2 SNP G5755e5, was previously shown to be associated with vascular disease, *e.g.*, MI and CAD. Individuals homozygous for the variant allele (GG) were at greater than 2-fold decreased odds of having vascular disease. Homozygous carriers of the variant allele of the G5755e9 SNP (CC) also showed a ~3-fold decreased odds of vascular disease. These two SNPs, G5755e5 and G5755e9, are in significant negative linkage disequilibrium with each other (D'=.49 (-), p=.04). The two SNPs together reveal distinct patterns of risk. Pattern 1 comprises two copies of the variant allele of G5755e9 (CC) in combination with two copies of the reference allele of G5755e5 (TT). Pattern 2 comprises two copies of the reference allele of G5755e9 (TT) and two copies of the variant allele of G5755e5 (GG) (see Table 2, below). Patterns 1 and 2 may independently influence risk of CAD. Individuals who have pattern 1 or pattern 2 are at ~3-fold decreased odds of vascular disease (odds ratio=0.32, p=.001) (see Table 3, below).

Table 2.

	G5755e9	G5755e5	CAD	controls	OR	P
	cc	gg/gt	0	0	-	-
1	cc	tt	2	6	0.38	ns
	tc	gg	5	3	1.89	ns
	tc	gt	25	29	0.98	ns
	tc	tt	38	40	1.08	ns
2	tt	gg	9	30	0.34	.01
	tt	gt	108	99	1.24	.31
	tt	tt	103	117	1.00	-

ns=non-significant; OR=odds ratio. All odds ratios are relative to last group in table.

Table 3.

	CAD	control
patterns 1 or 2	11	36
other	279	288

Odds ratio: 0.32 p=.001

ACE

A SNP in the ACE gene, identified herein as G765u2, has been identified which is also associated with a decreased risk of vascular disease, *e.g.*, MI and CAD. The G765u2 SNP is a change from an adenine (A) to a guanine (G) at nucleotide residue 86408 of the ACE reference sequence GI 13027555. This SNP is a "silent" variant. That is, it does not result in a change in the amino acid sequence of the ACE protein (see Table 4, below). Individuals with one copy of an A (the reference allele) and one copy of a G (the variant allele) at nucleotide residue 86408 of the ACE reference sequence GI 13027555 (AG genotype) are at a decreased risk for CAD and/or MI (CAD odds ratio:0.71; MI odds ratio:.66) (see Table 5, below).

An insertion/deletion polymorphism in the ACE gene was previously associated with vascular disease, *e.g.*, associated with a decreased risk for MI. The G765u2 SNP may be found to be in linkage disequilibrium with the previously identified insertion/deletion polymorphism. If these two polymorphisms are in linkage disequilibrium (LD), the G765u2 SNP would act as a marker for the insertion/deletion

polymorphism. Regardless of LD between these two polymorphisms, the G765u2 SNP represents a novel association with vascular disease.

Table 4.

1	2	3	4	5	6	7	8	9	10
Gene	**PolyID**	**variant freq.**	**Type of var.**	**Geno-types**	**Ref.**	**Var.**	**Genbank Accession No./nt position**	**Flanking sequence**	**SEQ ID NO:**
ACE	G765u2		silent	GG AG AA	A	G	GI: 13027555/ nt 86408	GAATGTG ATGGCCA CgTCCCG GAAATAT GAA	9

Table 5.

Gene	PolyID	Geno-type	Controls	CAD cases	MI cases	CAD Odds Ratio	I Odds Ratio
ACE	G765u2	GG	78	78	43	1.05 (.71, 1.56)	1.05 (.66, 1.68)
		AG	185	124	64	0.71 (.51, .98)	0.66 (.44, .95)
		AA	137	130	72	1.00	1.00

FGB

Two SNPs in the FGB gene, identified herein as FGBu1 and FGBu4, have been identified which are associated with decreased risk of vascular disease, *e.g.*, CAD and/or MI. The first SNP, FGBu1, is a change from a cytidine (C) to a thymidine (T) at nucleotide residue 5118 of the FGB reference sequence GI 182597. This SNP is a silent variant. The second SNP, FGBu4, is a change from a guanine (G) to an adenine (A) at nucleotide residue 8059 in the reference sequence GI 182597. This polymorphism is a missense variation which results in a change from an arginine (R) to a lysine (K) in the amino acid sequence of FGB (SEQ ID NO:6) at amino acid residue 478 (see Table 6, below).

For the FGBu1 SNP, individuals with two copies of a T (the variant allele) at nucleotide residue 5119 of the FGB reference sequence GI 182597 are at a decreased risk for CAD and MI (CAD odds ratio: 0.28; MI odds ratio: 0.43). Individuals with one

copy of a T and one copy of a C (the reference allele) at nucleotide residue 5119 of the FGB reference sequence GI 182597 are also at a decreased risk for CAD and MI (CAD odds ratio: 0.66; MI odds ratio: 0.72) (see Table 7, below).

For the FGBu4 SNP, individuals with two copies of an A (the variant allele) at nucleotide residue 8059 of the FGB reference sequence GI 182597 are at a decreased risk for CAD and MI (CAD odds ratio: 0.28; MI odds ratio: 0.43). Individuals with one copy of an A and one copy of a G (the reference allele) at nucleotide residue 5119 of the FGB reference sequence GI 182597 are also at a decreased risk for CAD and MI (CAD odds ratio: 0.61; MI odds ratio: 0.66) (see Table 7).

Two variants in the promoter region of the FGB gene at nucleotide residues –455 and –655, have been previously associated with vascular disease, *e.g.*, CAD and MI. The FGBu1 and FGBu4 SNPs may be found to be in linkage disequilibrium with these two previously identified SNPs. If these four SNPs are in linkage disequilibrium (LD), the FGBu1 and FGBu4 SNPs would act as markers for the previously identified SNPs. Regardless of LD, the FGBu1 and FGBu4 SNPs represent novel associations with vascular disease.

Table 6.

1	2	3	4	5	6	7	8	9	10
Gene	**PolyID**	**variant freq.**	**Type of var.**	**Geno-types**	**Ref.**	**Var.**	**Genbank Accession No./nt position**	**Flanking sequence**	**SEQ ID NO:**
FGB	FGBu1		silent	TT CT CC	C	T	GI: 182597/ nt 5119	TGAGACTG TGAATAGtA ATATCCCA ACTAAC	10
FGB	FGBu4		Missense (R/K)	AA AG GG	G	A	GI: 182597/ nt 8059	CATGGTAC TCAATGAa GAAGATGA GTATGAA	11

Table 7.

Gene	PolyID	Geno-type	Controls	CAD cases	MI cases	CAD Odds Ratio	MI Odds Ratio
FGB	FGBu1	TT	19	5	4	0.28 (.10, 76)	0.43 (.14, 1.28)
		CT	133	83	47	0.66 (.48, .92)	0.72 (.48, 1.07)
		CC	254	240	125	1.00	1.00
FGB	FGBu4	AA	19	5	4	0.28 (.10, .76)	0.43 (.14, 1.30)
		AG	137	78	44	0.61 (.44, .84)	0.66 (.44, .99)
		GG	255	239	124	1.00	1.00

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. A method for diagnosing or aiding in the diagnosis of a vascular disease or disorder in a subject comprising the steps of determining the THBS2, ACE, and FGB genetic profile of the subject, thereby diagnosing or aiding in the diagnosis of a vascular disease or disorder.

2. The method of claim 1, wherein determining the subject's THBS2 genetic profile comprises determining the identity of the nucleotide present at nucleotide position 3949 and/or 4476 of SEQ ID NO:1, or the complement thereof.

3. The method of claim 1, wherein determining the subject's ACE genetic profile comprises determining the identity of the nucleotide present at nucleotide position 86408 of SEQ ID NO:3, or the complement thereof.

4. The method of claim 1, wherein determining the subject's FGB genetic profile comprises determining the identity of the nucleotide present at nucleotide position 5119 and/or 8059 of SEQ ID NO:5, or the complement thereof.

5. The method of claim 1, wherein determining the subject's FGB genetic profile comprises determining the identity of the amino acid present at amino acid residue 478 of SEQ ID NO:6.

6. The method of claim 1, wherein the vascular disease is myocardial infarction.

7. The method of claim 1, wherein the vascular disease is coronary artery disease.

8. A method for predicting the likelihood that a subject will or will not develop a vascular disease or disorder comprising the steps of determining the THBS2,

ACE, and FGB genetic profile of the subject, thereby predicting the likelihood that a subject will or will not develop a vascular disease or disorder.

9. The method of claim 8, wherein determining the subject's THBS2 genetic profile comprises determining the identity of the nucleotide present at nucleotide position 3949 and/or 4476 of SEQ ID NO:1, or the complement thereof.

10. The method of claim 8, wherein determining the subject's ACE genetic profile comprises determining the identity of the nucleotide present at nucleotide position 86408 of SEQ ID NO:3, or the complement thereof.

11. The method of claim 8, wherein determining the subject's FGB genetic profile comprises determining the identity of the nucleotide present at nucleotide position 5119 and/or 8059 of SEQ ID NO:5, or the complement thereof.

12. The method of claim 8, wherein determining the subject's FGB genetic profile comprises determining the identity of the amino acid present at amino acid residue 478 of SEQ ID NO:6.

13. The method of claim 8, wherein the vascular disease is myocardial infarction.

14. The method of claim 8, wherein the vascular disease is coronary artery disease.

15. A method of diagnosing or aiding in the diagnosis of a vascular disease in a subject comprising the steps of determining the nucleotide present at nucleotide position 3949 and/or 4476 of SEQ ID NO:1, wherein the presence of two copies of a cytidine allele at nucleotide position 3949 of SEQ ID NO:1 together with two copies of a thymidine allele at nucleotide position 4476 of SEQ ID NO:1, or the complements thereof, and/or the presence of two copies of a thymidine allele at nucleotide position

3949 of SEQ ID NO:1 together with two copies of a guanine allele at nucleotide position 4476 of SEQ ID NO:1, or the complements thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

16. The method of claim 15, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

17. A method of diagnosing or aiding in the diagnosis of a vascular disease in a subject comprising the steps of determining the nucleotide present at nucleotide position 86408 of SEQ ID NO:3, wherein the presence of one copy of an adenine allele and one copy of a guanine allele at nucleotide position 3949 of SEQ ID NO:1, or the complement thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

18. The method of claim 17, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

19. A method of diagnosing or aiding in the diagnosis of a vascular disease in a subject comprising the steps of determining the nucleotide present at nucleotide position 5119 of SEQ ID NO:5, wherein the presence of two copies of a thymidine allele at position 5119 or the presence of one copy of a thymidine allele and one copy of a cytidine allele at position 5119, or the complements thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

20. The method of claim 19, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

21. A method of diagnosing or aiding in the diagnosis of a vascular disease in a subject comprising the steps of determining the nucleotide present at nucleotide

position 8059 of SEQ ID NO:5, wherein the presence of two copies of an adenine allele at position 8059 or the presence of one copy of an adenine allele and one copy of a guanine allele at position 8059, or the complements thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

22. The method of claim 21, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

23. The method of any one of claims 15, 17, 19, or 21, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary artery disease, myocardial infarction, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

24. The method of claim 23, wherein the vascular disease is myocardial infarction.

25. The method of claim 23, wherein the vascular disease is coronary artery disease.

26. A method for predicting the likelihood that a subject will or will not develop a vascular disease, comprising the steps of determining the nucleotide present at nucleotide position 3949 and/or 4476 of SEQ ID NO:1, wherein the presence of two copies of a cytidine allele at nucleotide position 3949 of SEQ ID NO:1 together with two copies of a thymidine allele at nucleotide position 4476 of SEQ ID NO:1, or the complements thereof, and/or the presence of two copies of a thymidine allele at nucleotide position 3949 of SEQ ID NO:1 together with two copies of a guanine allele at nucleotide position 4476 of SEQ ID NO:1, or the complements thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

27. The method of claim 26, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

28. A method for predicting the likelihood that a subject will or will not develop a vascular disease, comprising the steps of determining the nucleotide present at nucleotide position 86408 of SEQ ID NO:3, wherein the presence of one copy of an adenine allele and one copy of a guanine allele at nucleotide position 3949 of SEQ ID NO:1, or the complement thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

29. The method of claim 28, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

30. A method for predicting the likelihood that a subject will or will not develop a vascular disease, comprising the steps of determining the nucleotide present at nucleotide position 5119 of SEQ ID NO:5, wherein the presence of two copies of a thymidine allele at position 5119 or the presence of one copy of a thymidine allele and one copy of a cytidine allele at position 5119, or the complements thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject having any other combination of these alleles.

31. The method of claim 30, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

32. A method for predicting the likelihood that a subject will or will not develop a vascular disease, comprising the steps of determining the nucleotide present at nucleotide position 8059 of SEQ ID NO:5, wherein the presence of two copies of an adenine allele at position 8059 or the presence of one copy of an adenine allele and one copy of a guanine allele at position 8059, or the complements thereof, is indicative of decreased likelihood of a vascular disease in the subject as compared with a subject

having any other combination of these alleles.

33. The method of claim 32, wherein determining the identity of the nucleotides is by obtaining a nucleic acid sample from the subject.

34. The method of any one of claims 26, 28, 30, or 32, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary artery disease, myocardial infarction, ischemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

35. The method of claim 34, wherein the vascular disease is myocardial infarction.

36. The method of claim 34, wherein the vascular disease is coronary artery disease.

37. A computer readable medium for storing instructions for performing a computer implemented method for determining whether or not a subject has a predisposition to a vascular disease or disorder, said instructions comprising the functionality of:

obtaining information from the subject indicative of the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, and

based on the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, determining whether or not the subject has a predisposition to a vascular disease or disorder.

38. A computer readable medium for storing instructions for performing a computer implemented method for identifying a predisposition to a vascular disease or disorder, said instructions comprising the functionality of:

obtaining information regarding the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, and

based on the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, identifying a predisposition to a vascular disease or disorder.

39. An electronic system comprising a processor for determining whether or not a subject has a predisposition to a vascular disease or disorder, said processor implementing the functionality of:

obtaining information from the subject indicative of the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, and

based on the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, determining whether or not the subject has the predisposition to a vascular disease or disorder.

40. An electronic system comprising a processor for performing a method for identifying a predisposition to a vascular disease or disorder in a subject, said processor implementing the functionality of:

obtaining information from the subject indicative of the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, and

based on the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, performing a method for identifying a predisposition to a vascular disease or disorder associated with the polymorphic region.

41. The electronic system of claims 39 or 40, wherein said processor further implements the functionality of receiving phenotypic information associated with the subject.

42. The electronic system of claims 39 or 40, wherein said processor further implements the functionality of acquiring from a network phenotypic information associated with the subject.

43. A network system for identifying a predisposition to a vascular disease or disorder in response to information submitted by an individual, said system comprising

means for:

receiving data from the individual regarding the presence or absence of the polymorphic region of a THBS2, ACE, and/or FGB gene, and

based on the presence or absence of the polymorphic region, determining whether or not the subject has the predisposition to the vascular disease or disorder associated with the polymorphic region.

44. A network system for identifying whether or not a subject has a predisposition to a vascular disease or disorder, said system comprising means for:

receiving information from the subject regarding the polymorphic region of a THBS2, ACE, and/or FGB gene,

receiving phenotypic information associated with the subject,

acquiring additional information from the network, and

based on one or more of the phenotypic information, the polymorphic region, and the acquired information, determining whether or not the subject has a pre-disposition to a vascular disease or disorder associated with a polymorphic region of a THBS2, ACE, and/or FGB gene.

45. The system of claims 43 and 44, wherein the network system comprises a server and a work station operatively connected to said server via the network.

46. A composition comprising an isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, or a portion thereof, wherein residue 3949 is a cytidine, or the complement thereof, in combination with an isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, or a portion thereof, wherein residue 4476 is a thymidine, or the complement thereof.

47. A composition comprising an isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, or a portion thereof, wherein residue 3949 is a thymidine, or the complement thereof, in combination with an isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, or a

portion thereof, wherein residue 4476 is a guanine, or the complement thereof.

48. A composition comprising an isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3, or a portion thereof, wherein residue 86408 is a guanine, or the complement thereof.

49. A composition comprising an isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:5, or a portion thereof, wherein residue 8059 is an adenine, or the complement thereof.

50. A kit comprising probes or primers which are capable of hybridizing to the nucleic acid molecules of any of claims 46-49.

51. The kit of claim 50, wherein the probes or primers comprise a nucleotide sequence from about 15 to about 30 nucleotides.

52. The kit of claim 50, wherein the probes or primers are labeled.

53. A method for determining the identity of one or more allelic variants of a polymorphic region of a THBS2, ACE, and/or a FGB gene in a nucleic acid obtained from a subject, comprising contacting a sample nucleic acid from the subject with a probe or primer having a sequence which is complementary to a THBS2, ACE, and/or a FGB gene sequence, wherein the sample comprises a THBS2, ACE, and/or a FGB gene, thereby determining the identity of one or more of the allelic variants.

54. The method of claim 53, wherein the probes or primers are capable of hybridizing to an allelic variant of a polymorphic region of a THBS2, ACE, or FGB gene.

55. The method of claim 54, wherein determining the identity of the allelic variant comprises determining the identity of at least one nucleotide of the polymorphic region of a THBS2, ACE, or FGB gene.

56. The method of claim 55, wherein determining the identity of the allelic variant consists of determining the nucleotide content of the polymorphic region.

57. The method of claim 55, wherein determining the nucleotide content comprises sequencing the nucleotide sequence.

58. The method of claim 55, wherein determining the identity of the allelic variant comprises performing a restriction enzyme site analysis.

59. The method of claim 55, wherein determining the identity of the allelic variant is carried out by single-stranded conformation polymorphism.

60. The method of claim 55, wherein determining the identity of the allelic variant is carried out by allele specific hybridization.

61. The method of claim 55, wherein determining the identity of the allelic variant is carried out by primer specific extension.

62. The method of claim 55, wherein determining the identity of the allelic variant is carried out by an oligonucleotide ligation assay.

63. The method of claim 55, wherein the probe or primer comprises a nucleotide sequence from about 15 to about 30 nucleotides.

64. An Internet-based method for assessing a subject's risk for vascular disease, the method comprising:

a) analyzing biological information from a subject indicative of the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB;

b) providing results of the analysis to the subject via the Internet, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

65. A method of assessing a subject's risk for vascular disease, the method comprising:

a) obtaining biological information from the individual;

b) analyzing the information to obtain the subject's THBS2, ACE, and/or FGB genetic profile;

c) representing the THBS2, ACE, and/or FGB genetic profile information as digital genetic profile data;

d) electronically processing the THBS2, ACE, and/or FGB digital genetic profile data to generate a risk assessment report for vascular disease, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease; and

e) displaying the risk assessment report on an output device.

66. A method of assessing a subject's risk for vascular disease, the method comprising:

a) obtaining the subject's THBS2, ACE, and/or FGB genetic profile information as digital genetic profile data;

b) electronically processing the THBS2, ACE, and/or FGB digital genetic profile data to generate a risk assessment report for vascular disease, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease; and

c) displaying the risk assessment report on an output device.

67. The method of claims 65 or 66, further comprising the step of using the risk assessment report to provide medical advice.

68. The method of claims 65 or 66, wherein additional health information is provided.

69. The method of claim 68, wherein the additional health information comprises information regarding one or more of age, sex, ethnic origin, diet, sibling health, parental health, clinical symptoms, personal health history, blood test data, weight, and alcohol use, drug use, nicotine use, and blood pressure.

70. The method of claim 66, wherein the THBS2, ACE, and/or FGB digital genetic profile data are transmitted via a communications network to a medical information system for processing.

71. The method of claim 70, wherein the communications network is the Internet.

72. A medical information system for assessing a subject's risk for vascular disease comprising:

a) means for obtaining biological information from the individual to obtain a THBS2, ACE, and/or FGB genetic profile;

b) means for representing the THBS2, ACE, and/or FGB genetic profile as digital molecular data;

c) means for electronically processing the THBS2, ACE, and/or FGB digital genetic profile to generate a risk assessment report for vascular disease; and

d) means for displaying the risk assessment report on an output device, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

73. A medical information system for assessing a subject's risk for vascular disease comprising:

a) means for representing the subject's THBS2, ACE, and/or FGB genetic profile data as digital molecular data;

b) means for electronically processing the THBS2, ACE, and/or FGB digital genetic profile to generate a risk assessment report for vascular disease; and

c) means for displaying the risk assessment report on an output device, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

74. A computerized method of providing medical advice to a subject comprising:

a) analyzing biological information from a subject to determine the subject's THBS2, ACE, and/or FGB genetic profile;

b) based on the subject's THBS2, ACE, and/or FGB genetic profile, determining the subject's risk for vascular disease;

c) based on the subject's risk for vascular disease, electronically providing medical advice to the subject.

75. A computerized method of providing medical advice to a subject comprising:

a) based on the subject's THBS2, ACE, and/or FGB genetic profile, determining the subject's risk for vascular disease;

b) based on the subject's risk for vascular disease, electronically providing medical advice to the subject.

76. The method of any of claims 74 or 75, wherein the medical advice comprises one or more of the group consisting of further diagnostic evaluation, administration of medication, or lifestyle change.

77. The method of claims 74 or 75, wherein additional health information is obtained from the subject.

78. The method of claim 77, wherein the additional health information comprises information regarding one or more of age, sex, ethnic origin, diet, sibling health, parental health, clinical symptoms, personal health history, blood test data, weight, and alcohol use, drug use, nicotine use, and blood pressure.

79. A method for self-assessing risk for a vascular disease comprising

a) providing biological information for genetic analysis;

b) accessing an electronic output device displaying results of the genetic analysis, thereby self-assessing risk for a vascular disease, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

80. A method for self-assessing risk for a vascular disease comprising accessing an electronic output device displaying results of a genetic analysis of a biological sample, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease, thereby self-assessing risk for a vascular disease.

81. A method of self-assessing risk for vascular disease, the method comprising

a) providing biological information;

b) accessing THBS2, ACE, and/or FGB digital genetic profile data obtained from the biological information, the THBS2, ACE, and/or FGB digital genetic profile data being displayed via an output device, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

82. A method of self-assessing risk for vascular disease, the method comprising accessing THBS2, ACE, and/or FGB digital genetic profile data obtained from biological information, the THBS2, ACE, and/or FGB digital genetic profile data

being displayed via an output device, wherein the presence of a polymorphic region of THBS2, ACE, and/or FGB indicates a decreased risk for vascular disease.

83. The method of claims 81 or 82, wherein the electronic output device is accessed via the Internet.

84. The method of claims 81 or 82, wherein additional health information is provided.

85. The method of claim 84, wherein the additional health information comprises information regarding one or more of age, sex, ethnic origin, diet, sibling health, parental health, clinical symptoms, personal health history, blood test data, weight, and alcohol use, drug use, nicotine use, and blood pressure.

86. The method of any of claims 79, 80, 81, or 82, wherein the biological information is obtained from a sample from an individual at a laboratory company.

87. The method of claim 86, wherein the laboratory company processes the biological sample to obtain THBS2, ACE, and/or FGB genetic profile data, represents at least some of the THBS2, ACE, and/or FGB genetic profile data as digital genetic profile data, and transmits the THBS2, ACE, and/or FGB digital genetic profile data via a communications network to a medical information system for processing.

88. The method of any of claims 79, 80, 81, or 82, wherein the biological information is obtained from a sample from an individual at a draw station, wherein the draw station processes the biological sample to obtain THBS2, ACE, and/or FGB genetic profile data, and transfers the data to a laboratory company.

89. The method of claim 88, wherein the laboratory company represents at least some of the THBS2, ACE, and/or FGB genetic profile data as digital genetic profile data, and transmits the THBS2, ACE, and/or FGB digital genetic profile data via

a communications network to a medical information system for processing.

90. A method for a health care provider to generate a personal health assessment report for an individual, the method comprising counseling the individual to provide a biological sample; authorizing a draw station to take a biological sample from the individual and transmit molecular information from the sample to a laboratory company, wherein the molecular information comprises the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB; requesting the laboratory company to provide digital molecular data corresponding to the molecular information to a medical information system to electronically process the digital molecular data and digital health data obtained from the individual to generate a health assessment report; receiving the health assessment report from the medical information system; and providing the health assessment report to the individual.

91. A method for a health care provider to generate a personal health assessment report for an individual, the method comprising requesting a laboratory company to provide digital molecular data corresponding to the molecular information derived from a biological sample from the individual to a medical information system to electronically process the digital molecular data and digital health data obtained to generate a health assessment report; receiving the health assessment report from the medical information system; and providing the health assessment report to the individual.

92. A method of assessing the health of an individual, the method comprising:

obtaining health information from the individual using an input device; representing at least some of the health information as digital health data; obtaining biological information from the individual, wherein the information comprises the presence or absence of a polymorphic region of THBS2, ACE, and/or FGB; representing at least some of the information as digital molecular data; electronically processing the digital molecular data and digital health data to generate a health assessment report; and

displaying the health assessment report on an output device.

93. The method of claim 92, wherein electronically processing the digital molecular data and digital health data to generate a health assessment report comprises using the digital molecular data and digital health data as inputs for an algorithm or a rule-based system that determines whether the individual is at risk for a specific disorder.

94. The method of claim 92, wherein the individual has or is at risk of developing vascular disease, and wherein electronically processing the digital molecular data and digital health data to generate a health assessment report comprises using the digital molecular data and digital health data as inputs for an algorithm or a rule-based system that determines the individual's prognosis.

95. The method of claim 92, wherein electronically processing the digital molecular data and digital health data comprises using the digital molecular data and digital health data as inputs for an algorithm or a rule-based system based on one or more databases comprising stored digital molecular data and/or digital health data relating to one or more disorders.

96. The method of claim 92, wherein electronically processing the digital molecular data and digital health data comprises using the digital molecular data and digital health data as inputs for an algorithm or a rule-based system based on one or more databases comprising (i) stored digital molecular data and/or digital health data from a plurality of healthy individuals, and (ii) stored digital molecular data and/or digital health data from one or more pluralities of unhealthy individuals, each plurality of individuals having a specific disorder.

97. The method of either of claims 95 or 96, wherein at least one of the databases is a public database.

98. The method of claim 92, wherein the digital health data and digital molecular data are transmitted via a communications network to a medical information system for processing.

99. The method of claim 98, wherein the communications network is the Internet.

100. The method of claim 98, wherein the input device is a keyboard, touch screen, hand-held device, telephone, wireless input device, or interactive page on a website.

101. The method of claim 92, wherein the health assessment report comprises a digital molecular profile of the individual.

102. The method of claim 92, wherein the health assessment report comprises a digital health profile of the individual.

103. The method of claim 92, wherein the molecular data comprises nucleic acid sequence data, and the molecular profile comprises a genetic profile.

104. The method of claim 92, wherein the molecular data comprises protein sequence data, and the molecular profile comprises a proteomic profile.

105. The method of claim 92, wherein the molecular data comprises information regarding one or more of the absence, presence, or level, of one or more specific proteins, polypeptides, chemicals, cells, organisms, or compounds in the individual's biological sample.

106. The method of claim 92, wherein the health information comprises information relating to one or more of age, sex, ethnic origin, diet, sibling health, parental health, clinical symptoms, personal health history, blood test data, weight, and

alcohol use, drug use, nicotine use, and blood pressure.

107. The method of claim 92, wherein the health information comprises current and historical health information.

108. The method of claim 92, further comprising obtaining a second set of biological information at a time after obtaining the first set of biological information; processing the second set of biological information to obtain a second set of information; representing at least some of the second set of information as digital second molecular data; and processing the molecular data and second molecular data to generate a health assessment report.

109. The method of claim 108, further comprising obtaining second health information at a time after obtaining the health information; representing at least some of the second health information as digital second health data and processing the molecular data, health data, second molecular data, and second health data to generate a health assessment report.

110. The method of claim 92, wherein the health assessment report provides information about the individual's predisposition for vascular disease and options for risk reduction.

111. The method of claim 110, wherein the options for risk reduction comprise one or more of diet, exercise, one or more vitamins, one or more drugs, cessation of nicotine use, and cessation of alcohol use.

112. The method of claim 85, wherein the health assessment report provides information about treatment options for a particular disorder.

113. The method of claim 107, wherein the treatment options comprise one or more of diet, one or more drugs, physical therapy, and surgery.

114. The method of claim 85, wherein the health assessment report provides information about the efficacy of a particular treatment regimen and options for therapy adjustment.

115. The method of claim 85, further comprising storing the molecular data.

116. The method of claim 115, further comprising building a database of stored molecular data from a plurality of individuals.

117. The method of claim 92, further comprising storing the molecular data and health data.

118. The method of claim 117, further comprising building a database of stored molecular data and health data from a plurality of individuals.

119. The method of claim 118, further comprising building a database of stored digital molecular data and/or digital health data from a plurality of healthy individuals, and stored digital molecular data and/or digital health data from one or more pluralities of unhealthy individuals, each plurality of individuals having a specific disorder.

FIGURE 1

```
1    acggcatcca gtacagaggg gctggacttg gacccctgca gcagccctgc acaggagaag
61   cggcatataa agccgcgctg cccgggagcc gctcggccac gtccaccgga gcatcctgca
121  ctgcagggcc ggtctctcgc tccagcagag cctgcgcctt tctgactcgg tccggaacac
181  tgaaaccagt catcactgca tcttttggc aaaccaggag ctcagctgca ggaggcagga
241  tggtctggag gctggtcctg ctggctctgt gggtgtggcc cagcacgcaa gctggtcacc
301  aggacaaaga cacgaccttc gacctttca gtatcagcaa catcaaccgc aagaccattg
361  gcgccaagca gttccgcggg cccgaccccg gcgtgccggc ttaccgcttc gtgcgctttg
421  actacatccc accggtgaac gcagatgacc tcagcaagat caccaagatc atgcggcaga
481  aggagggctt cttcctcacg gcccagctca agcaggacgg caagtccagg ggcacgctgt
541  tggctctgga gggccccggt ctctcccaga ggcagttcga gatcgtctcc aacggccccg
601  cggacacgct ggatctcacc tactggattg acggcacccg gcatgtggtc tccctggagg
661  acgtcggcct ggctgactcg cagtggaaga acgtcaccgt gcaggtggct ggcgagacct
721  acagcttgca cgtgggctgc gacctcatag gaccagttgc tctggacgag cccttctacg
781  agcacctgca ggcggaaaag agccggatgt acgtggccaa aggctctgcc agagagagtc
841  acttcagggg tttgcttcag aacgtccacc tagtgtttga aaactctgtg gaagatattc
901  taagcaagaa gggttgccag caaggccagg gagctgagat caacgccatc agtgagaaca
961  cagagacgct gcgcctgggt ccgcatgtca ccaccgagta cgtgggcccc agctcggaga
1021 ggaggcccga ggtgtgcgaa cgctcgtgcg aggagctggg aaacatggtc caggagctct
1081 cggggctcca cgtcctcgtg aaccagctca gcgagaacct caagagagtg tcgaatgata
1141 accagtttct ctgggagctc attggtggcc ctcctaagac aaggaacatg tcagcttgct
1201 ggcaggatgg ccggttcttt gcggaaaatg aaacgtgggt ggtggacagc tgcaccacgt
1261 gtacctgcaa gaaatttaaa accatttgcc accaaatcac ctgcccgcct gcaacctgcg
1321 ccagtccatc ctttgtggaa ggcgaatgct gcccttcctg cctccactcg gtggacggtg
1381 aggagggctg gtctccgtgg gcagagtgga cccagtgctc cgtgacgtgt ggctctggga
1441 cccagcagag aggccggtcc tgtgacgtca ccagcaacac ctgcttgggg ccctcgatcc
1501 agacacgggc ttgcagtctg agcaagtgtg acacccgcat ccggcaggac ggcggctgga
1561 gccactggtc accttggtct tcatgctctg tgacctgtgg agttggcaat atcacacgca
1621 tccgtctctg caactcccca gtgccccaga tgggggcaa gaattgcaaa gggagtggcc
1681 gggagaccaa agcctgccag ggcgccccat gcccaatcga tggccgctgg agcccctggt
1741 ccccgtggtc ggcctgcact gtcacctgtg ccggtgggat ccgggagcgc acccgggtct
1801 gcaacagccc tgagcctcag tacggaggga aggcctgcgt gggggatgtg caggagcgtc
1861 agatgtgcaa caagaggagc tgccccgtgg atggctgttt atccaacccc tgcttcccgg
1921 gagcccagtg cagcagcttc cccgatgggt cctggtcatg cggcttctgc cctgtgggct
1981 tcttgggcaa tggcacccac tgtgaggacc tggacgagtg tgccctggtc cccgacatct
2041 gcttctccac cagcaaggtg cctcgctgtg tcaacactca gcctggcttc cactgcctgc
2101 cctgcccgcc ccgatacaga gggaaccagc ccgtcggggt cggcctggaa gcagccaaga
2161 cggaaaagca agtgtgtgag cccgaaaacc catgcaagga caagacacac aactgccaca
2221 agcacgcgga gtgcatctac ctgggtcact tcagcgaccc catgtacaag tgcgagtgcc
2281 agacaggcta cgcgggcgac gggctcatct gcggggagga ctcggacctg gacggctggc
2341 ccaacctcaa tctggtctgc gccaccaacg ccacctacca ctgcatcaag gataactgcc
2401 cccatctgcc aaattctggg caggaagact ttgacaagga cgggattggc gatgcctgtg
2461 atgatgacga tgacaatgac ggtgtgaccg atgagaagga caactgccag ctcctcttca
2521 atccccgcca ggctgactat gacaaggatg aggttgggga ccgctgtgac aactgccctt
2581 acgtgcacaa ccctgcccag atcgacacag acaacaatgg agagggtgac gcctgctccg
2641 tggacattga tggggacgat gtcttcaatg aacgagacaa ttgtccctac gtctacaaca
2701 ctgaccagag ggacacggat ggtgacggtg tgggggatca ctgtgacaac tgccccctgg
2761 tgcacaaccc tgaccagacc gacgtggaca atgaccttgt tggggaccag tgtgacaaca
2821 acgaggacat agatgacgac ggccaccaga acaaccagga caactgcccc tacatctcca
2881 acgccaacca ggctgaccat gacagagacg gccagggcga cgcctgtgac cctgatgatg
2941 acaacgatgg cgtccccgat gacagggaca actgccggct tgtgttcaac ccagaccagg
3001 aggacttgga cggtgatgga cggggtgata tttgtaaaga tgattttgac aatgacaaca
3061 tcccagatat tgatgatgtg tgtcctgaaa acaatgccat cagtgagaca gacttcagga
```

FIGURE 1 (continued)

3121 acttccagat ggtccccttg gatcccaaag ggaccaccca aattgatccc aactgggtca
3181 ttcgccatca aggcaaggag ctggttcaga cagccaactc ggaccccggc atcgctgtag
3241 gttttgacga gtttgggtct gtggacttca gtggcacatt ctacgtaaac actgaccggg
3301 acgacgacta tgctggcttc gtctttggtt accagtcaag cagccgcttc tatgtggtga
3361 tgtggaagca ggtgacgcag acctactggg aggaccagcc cacgcgggcc tatggctact
3421 ccggcgtgtc cctcaaggtg gtgaactcca ccacggggac gggcgagcac ctgaggaacg
3481 cgctgtggca cacggggaac acgccggggc aggtgcgaac cttatggcac gaccccagga
3541 acattggctg gaaggactac acggcctata ggtggcacct gactcacagg cccaagaccg
3601 gctacatcag agtcttagtg catgaaggaa aacaggtcat ggcagactca ggacctatct
3661 atgaccaaac ctacgctggc gggcggctgg gtctatttgt cttctctcaa gaaatggtct
3721 atttctcaga cctcaagtac gaatgcagag atatttaaac aagatttgct gcatttccgg
3781 caatgccctg tgcatgccat ggtccctaga cacctcagtt cattgtggtc cttgcggctt
3841 ctctctctag cagcacctcc tgtcccttga ccttaactct gatggttctt cacctcctgc
3901 cagcaacccc aaacccaagt gccttcagag gataaatatc aatggaactc agagatgaac
3961 atctaaccca ctagaggaaa ccagtttggt gatatatgag actttatgtg gagtgaaaat
4021 tgggcatgcc attacattgc tttttcttgt ttgtttaaaa agaatgacgt ttacatataa
4081 aatgtaatta cttattgtat ttatgtgtat atggagttga agggaatact gtgcataagc
4141 cattatgata aattaagcat gaaaaatatt gctgaactac ttttggtgct taaagttgtc
4201 actattcttg aattagagtt gctctacaat gacacacaaa tcccgctaaa taaattataa
4261 acaagggtca attcaaattt gaagtaatgt tttagtaagg agagattaga agacaacagg
4321 catagcaaat gacataagct accgattaac taatcggaac atgtaaaaca gttacaaaaa
4381 taaacgaact ctcctcttgt cctacaatga aagccctcat gtgcagtaga gatgcagttt
4441 catcaaagaa caaacatcct tgcaaatggg tgtgacgcgg ttccagatgt ggatttggca
4501 aaacctcatt taagtaaaag gttagcagag caaagtgcgg tgctttagct gctgcttgtg
4561 ccgttgtggc gtcggggagg ctcctgcctg agcttccttc cccagctttg ctgcctgaga
4621 ggaaccagag cagacgcaca ggccggaaaa ggcgcatcta acgcgtatct aggctttggt
4681 aactgcggac aagttgcttt tacctgattt gatgatacat ttcattaagg ttccagttat
4741 aaatattttg ttaatattta ttaagtgact atagaatgca actccattta ccagtaactt
4801 attttaaata tgcctagtaa cacatatgta gtataatttc tagaaacaaa catctaataa
4861 gtatataatc ctgtgaaaat atgaggcttg ataatattag gttgtcacga tgaagcatgc
4921 tagaagctgt aacagaatac atagagaata atgaggagtt tatgatggaa ccttaatata
4981 taatgttgcc agcgatttta gttcaatatt tgttactgtt atctatctgc tgtatatgga
5041 attcttttaa ttcaaacgct gaaaacgaat cagcatttag tcttgccagg cacacccaat
5101 aatcagtcat gtgtaatatg cacaagtttg tttttgtttt tgtttttttt gttggttggt
5161 ttttttgctt taagttgcat gatctttctg caggaaatag tcactcatcc cactccacat
5221 aaggggttta gtaagagaag tctgtctgtc tgatgatgga taggggggcaa atctttttcc
5281 cctttctgtt aatagtcatc acatttctat gccaaacagg aacgatccat aactttagtc
5341 ttaatgtaca cattgcattt tgataaaatt aattttgttg tttcctttga ggttgatcgt
5401 tgtgttgttt tgctgcactt tttacttttt tgcgtgtgga gctgtattcc cgagacaacg
5461 aagcgttggg atacttcatt aaatgtagcg actgtcaaca gcgtgcaggt tttctgtttc
5521 tgtgttgtgg ggtcaaccgt acaatggtgt gggaatgacg atgatgtgaa tatttagaat
5581 gtaccatatt ttttgtaaat tatttatgtt tttctaaaca aatttatcgt ataggttgat
5641 gaaacgtcat gtgttttgcc aaagactgta aatatttatt tatgtgttca catggtcaaa
5701 atttcaccac tgaaaccctg cacttagcta gaacctcatt tttaaagatt aacaacagga
5761 aataaattgt aaaaaaggtt ttct

FIGURE 2

```
1    mvwrlvllal wvwpstqagh qdkdttfdlf sisninrkti gakqfrgpdp gvpayrfvrf
61   dyippvnadd lskitkimrq kegffltaql kqdgksrgtl lalegpglsq rqfeivsngp
121  adtldltywi dgtrhvvsle dvgladsqwk nvtvqvaget yslhvgcdli gpvaldepfy
181  ehlqaeksrm yvakgsares hfrgllqnvh lvfensvedi lskkgcqqgq gaeinaisen
241  tetlrlgphv tteyvgpsse rrpevcersc eelgnmvqel sglhvlvnql senlkrvsnd
301  nqflweligg ppktrnmsac wqdgrffaen etwvvdsctt ctckkfktic hqitcppatc
361  aspsfvegec cpsclhsvdg eegwspwaew tqcsvtcgsg tqqrgrscdv tsntclgpsi
421  qtracslskc dtrirqdggw shwspwsscs vtcgvgnitr irlcnspvpq mggknckgsg
481  retkacqgap cpidgrwspw spwsactvtc aggirertrv cnspepqygg kacvgdvqer
541  qmcnkrscpv dgclsnpcfp gaqcssfpdg swscgfcpvg flgngthced ldecalvpdi
601  cfstskvprc vntqpgfhcl pcppryrgnq pvgvgleaak tekqvcepen pckdkthnch
661  khaeciylgh fsdpmykcec qtgyagdgli cgedsdldgw pnlnlvcatn atyhcikdnc
721  phlpnsgqed fdkdgigdac dddddndgvt dekdncqllf nprqadydkd evgdrcdncp
781  yvhnpaqidt dnngegdacs vdidgddvfn erdncpyvyn tdqrdtdgdg vgdhcdncpl
841  vhnpdqtdvd ndlvgdqcdn nedidddghq nnqdncpyis nanqadhdrd gqgdacdpdd
901  dndgvpddrd ncrlvfnpdq edldgdgrgd ickddfdndn ipdiddvcpe nnaisetdfr
961  nfqmvpldpk gttqidpnwv irhqgkelvq tansdpgiav gfdefgsvdf sgtfyvntdr
1021 dddyagfvfg yqsssrfyvv mwkqvtqtyw edqptraygy sgvslkvvns ttgtgehlrn
1081 alwhtgntpg qvrtlwhdpr nigwkdytay rwhlthrpkt gyirvlvheg kqvmadsgpi
1141 ydqtyaggrl glfvfsqemv yfsdlkyecr di
```

FIGURE 3

```
1    actttcttgc tgctctgcct gcatggacct gtgacaggca tcatctatng aatctcatgn
61   agcgtcgact ccctgcccca ggcattggtg agtggcaagt gagagctgct cacaggcatc
121  aaagggtcaa aatagcaccc aaggccgggc atgannncac gcctgtaatc ccagtacttt
181  gggaggctga ggtaggcgga tcacctgagg tcaggagctc aaagccagcc tggccaacat
241  ggagaaaccc cgtctctact aaaaatacaa aaattagctg ggtgtggtgg tgtgtgcctg
301  taatcccagc tacttggagg ctgaggcagg agaataactt gaacccggga ggcaaaagtt
361  gcagtgagca agattgcacc attgcactcg agcctgggga aaaagagcaa gactccatct
421  caaaaaaaaa aaaaaaaaat tgcacctaag cccaagccca gaaggtggcc ccgaacctgg
481  gccttccttt gaagccaggc ctggtatttg ccaagatcgt gcctcctgcc cccacatcac
541  caacgttaac tgccctccat tagctggtgc tgcctgggtg ctgcgtgggc gctggcgtgt
601  gaaatggcaa catgaggccg tgtctactga gccaggcatc ttagtgcttt acacagagaa
661  gtgaaacctc ctagcaaccc tgtgagttag gtaccatttg attcccattt ttcagataag
721  gaaaccaagg cacaaagtgg ttggtaggtg gactaaggtc acccaacaaa tggaggcaaa
781  acgggacacg tgtccccacc aggagccacc agctgtccct ccttctccat cccacagact
841  gcgtgaagcc ctgggctact ctggctggat ggtatcaagt gtcacattct gtggagccag
901  gatgttcctc caatccttta atcccaaatc ttttattgtg tttgttttgt tttgttttga
961  cagagtcttg ctctgtcgtc caggctggag tgctgtggtg caatctcagc tcactgcaac
1021 ctctgcctcc tgggttcaag tgattctcct gcctcatcct ccctagtagc tgggactaca
1081 ggtgcgcacc accaagcccg gctaattttt gtatttttg ttgagacggg gtttcaccat
1141 gttggccagg ctggttatga actcctgacc tcaaatgatc tgcctgcctc agcctcccaa
1201 agtgctggca ttacaggcat gagccaccat gcctggccta atcccaaatc ttaaacctac
1261 ttttcctggt tctgcctcta gtgaaatcaa agatccctcc gctgctgtcc ctatgggaat
1321 gtggacaggg atgacacctc tcgccagctc tgggtggccc ctctggtctg gtcctttgct
1381 gcccagctcc tcaagtcctc ctcaacaccc ctcttctcca gccctgctgc cactctcccc
1441 acctgagtac agaccctcat catttcccac tgggcccccc cttctcgtcc ccccagtccc
1501 tnctccctag ntccatttat ttattctnat tttgnnagac ggagttttgc tcttgttgcc
1561 caagctggag tgcaatggcg cgatctcagc tcactgcaac ctctgcctcc caggttcaag
1621 cgattttccn tgcctcagcc tcccaagtag ctgggattac aggcatgcac caccacaccc
1681 ggctaatttt ngtattttta gtagaggtgg ggtttctcca tgttggtcan ggctggtctc
1741 gaactcccga cctcaggtga tccacctgcc tcagcctgcc aaagtgctgg gattnacagg
1801 cgtgancanc tgntgnccct gcccctagtc catttttttt ttttttaatt gatcattctt
1861 gggtgtttct cgcagagggg gatttggcag ggtcacagga caatagtgga gggaaggtca
1921 gcagataaac aagtgaacaa aggtctctgg ttttcctagg cagaggaccc tgtggccttc
1981 cgcagtgttt gtgtccctgg gtacttgaga ttagggagtg gtgatgactc ttaaggagca
2041 tgctgccttc aagcatctgt ttaacaaagc acatcttgca ccgcccttaa tccattcaac
2101 cctgagtgga tacagcacat gtttcagaga gcacagggtt gggggtaagg tcacagatca
2161 acaggatccc aaggcagaag aatttttctt agtacagaac aaaatgaaaa gtctcccacg
2221 tctacctctt tctacacaga cacggcaacc atccgatttc tcaatctttt ccccaccctt
2281 ccccccttte tattccacaa aaccgccatt gtcatcatgg cccgttctca atgagctgtt
2341 gggtacacct cccagaccgg gtggtggccg ggcagagggg ctcctcactt cccagtaggg
2401 gcggccgggc agaggcgccc ctcancctcc cggacggggc ggctggccgg gcgggggget
2461 gacccccca cctccctccc ggacggggcg gccggccagg cagaggggct cctcacttcc
2521 cagtaggggc ggccgggcag aggcgccect cacctcccgg acagggcggc tggccgggca
2581 gggggctgat ccccccacct ccctcccgga cggggaggct ggccgggcgg ggggctgacc
2641 ccccctgacc cccccacctc cctcccggac ggggcggctg gctgggcggg gggctgaccc
2701 ccacacctcc ctcccggacg gggcagctgg ccgggcgggg ggctgacccc cccacctccc
2761 tcctggacgg agcggctggc cgggcagagg ggctcctcac ttcccagtag gggcggccgg
2821 gcagaggcgc ccctcacctc ccggacgggg cggctggccg ggncaggggg ctgatcgccc
2881 cacnctccct cccggacggg gaggctggcc gggcgggggg ctgacccccc cacctccctg
2941 ccggacgagg tggctgccgg gcagagacgc tcctcacttc ccagacgggg tggcttctgg
3001 acggatgggc tcctcacttc tcagacgggg cggttgccgg gcggagggtc tcctcacttc
```

FIGURE 3 (continued)

3061 tcagaggggg cggccgggca gagacgctcc tcacatcccg gacggggcgg cagggcagag
3121 gtgctcccca catctcagac gatgggcggc cgggcagaga cgctcctcac ttcccagatg
3181 ggatggctgc agggaagagg cgctcctcac ttcctagatg ggatggcggc caggcagaga
3241 cgctcctcac gtcccagacg atgggcggct gggcagagac gctcctcact tcccagacgg
3301 ggtggcggcc gggcagaggc tgcaatctcg gcactttggg aggccaaggc aggctgctgg
3361 gaggtgaagg ttgtagcgag ccgagatcac gccactgcac tccagcctgg gcaccattga
3421 gcacggagtg aacgagactc cgtctgcaat cccggcacct caggatgccg aggctggcgg
3481 atcactcgcg gttaggagct ggagaccagc ccggccaaca cagcaaaacc ccgtctccac
3541 caaaaaaata cgaaaaccag tcaggtgtgg cggcgcgcac ctgcaatcgc aggcagtcgg
3601 caggctgagg caggagaatc aggcagcagt accgtccagc ttcagctcgg catcagaggg
3661 angaccgtgg agagagggag agggagaccg tggggagagg gagagggaga gggagggaga
3721 gggagaggga gagggagagg gagagagcta gtccattttt atanncatgg ctctgaggat
3781 gctccaatct gaccacagct tctgttcact tagcgccctg ccccggtgtc tcatcacctt
3841 ggggataaaa tccacactct gagagtgagt gcaaagtcct tcctggcgtg gcctctgcct
3901 cctccatttc ctcaggcctt gctctaccac acaagtccct ctattagtgc ctcccgatgg
3961 ccagctccct ggcagggctg atagtgataa aatgttcagt agtcagcaat aaagaggcaa
4021 aaacagcaac actgatttat agatatgaaa tcaaccagat gtatcagcca agttgccagc
4081 attcctttct tggtcacaat atcttttatt ccaaacatcc tttctattta ttttttgtag
4141 ttaagatttt attacttctt ttgtaaaacg gtggtgagaa taagggacta tttttttaa
4201 atgtccttac ttggcaaact aaaaaatgtt tgcaactcta gtaagactcc actgggagct
4261 caggggctgg ggtgcctgat ctggccctaa agccccannn nnnnnnnnnn nnnnncngac
4321 actgacctgg tgcctgcgtg attcttcttc attgttccca gacaggggcc aggaaatgga
4381 atgaaagcag ccactgtctg aagagctgga gaccatcatc tgcctctgga agcccagaga
4441 acctcggctc agacagaagg acagagactg agggaaggga gagagactgt gacagagaag
4501 cagaggaggg tgacagagtc agggaggaac aaaacagcct gcagtgggag cagagacaga
4561 aatgtggggg acccacaggg anggggaggg agggaagggg agggacggag ggagggacaa
4621 ctgccgtcca agtggctgtg agagcctggg gctggggaga ggcaccctcc tcctgttggc
4681 ttctcataca gtctctatca ggggacccag gacacaagaa gcaaattcat ctttgggtcc
4741 acccttcagc tctagaagtc tagggcctgg ccaggtcctc tgaaggggtc tctggcccca
4801 ggcagcttcc tgcccccctg cgtggccccc caagcccttc tcaagagccc ttcagtaaat
4861 aaattcactg gaagacatca atgatgggtc ataggaggtg ttcctagggg agaaaatgac
4921 atttagtttt tagctcactt ggagccattt gctttaagaa gcacctcttg gccaggcgca
4981 gtgactcacg cctgtaatcc caacagtttg ggaggccgag gtgggcggat cacctgaggt
5041 caggagttca agaccagcct ggtcaacatg gtgaaaccct gtnctctact aaaaatacaa
5101 aaattagttg ggtgtggtgg catgtgcctg taatcccagc tgcttgggag gctgaggtga
5161 gagaatcgct tgaactcagg aggtggaggc tgcagtgagc tgagatcaca ccactgcatt
5221 ccagcctggg tgacacagtg agactatgtc taaaaaacaa aaaacaaaaa cgaagcacct
5281 cttgagagag ctttgcatgc agatgtcaca ggcaaaatgc ctcatgttgt ggggaatgtg
5341 gcagtggact tctggggcca ggctgaggcc cccacgtgct gtccccttg ggagcagggc
5401 agggtnggct gggttgacag atgtgcagag aaggagagaa gtgggggcaa tgccttctct
5461 ctgatggtca gctggcggga aaggctccag gggcttctga ggagagaccc taagaggact
5521 ctggatgggc ctcggctggg ggnaggctgg tgctgacctg gcacgtgggc actgtggggg
5581 tgcaatgggg gacagtgtgg gaggtgtcat ggggggtctt tacactgggt actaaagagc
5641 ttaggagcct gagctacctg gtgtggggca atcacatttg ctgctgtgtc caggcaagga
5701 ggagcaggtg ggcttggacc ccaaagtgac atgaagtggc caggactgag cgctgggcca
5761 acgtgagccc caggggtgc ttggatttat cagctgaaga agggcctatg gaaggggaga
5821 aaacaatcct ccagacagat gagagcactc agtcacatcc acctggacct gacatcacct
5881 cccaggctgt gcaggggcac acacagaact ggaatgggta atgacttatt agagcctcct
5941 tccaacagct gtgggccctg gaacctaaga acccaccttt ggagaacatt tttgggtctg
6001 agagctttga ggccaggcct tgagacaaac ccacccagaa gccctgagct acttttgctg
6061 taagagtagg nagggccctg cccctgctcc aggagccgcc ccttgatccc tggaagagaa
6121 ggaagcactg actgatccag aggggcgatg agtccagagg ctgagatcag ctggaaccgg

FIGURE 3 (continued)

6181 gccaccagct gagcgtgtga ttcctctggc tccaggccac ctctgagggt ctcccaatgg
6241 ctgggaggga ttagaggact gtggcttctt gggtgaggga gcctctgggt tacagtcacc
6301 gtgccctttg aatctaatct gctctcgaca agaatcagct gacctgatag gcatgctcag
6361 gcaggaggca natggccagg gccaagagaa gcataaggat ggagtagagg agggagcaga
6421 gcagggtcag gtgaaatctc tcactcagga caattgtgca gctgacttgc tgataccttg
6481 gaggagtcac gaagttcagt gtgcccagtt ggggcatctt ccaggctgtg gctccatcac
6541 cctgcagttt tggctcagaa cagccaccac atcctcactg ctctcctggg atgagaaggt
6601 gggtggagaa ccaggggctt ggtgcgtggc cagagctgga gcagcttcag gagccaggtg
6661 aaacctcaga gcctcacagg ctttgaacaa agtgcaggag tcccagccct gccatttatg
6721 gcgggttcct ggcggtgatc agtcattatc agcagtatgt tcactctgct gatatcaccc
6781 aaggccatga tcatccctta ccccaggcct tgcacacttg cttccctgtc cnnnnntaga
6841 cagctgggnn nngctgttcc tgctcttcgc tgcccctggc actcacagag cagctgactg
6901 actgacctac caacttaggg agaaaagtgg ggaggtgggg atggaggaca ctctggggat
6961 acagcctctt cgtgatttcc ttcctggagg ggagaggcgg ccagggcaat gcaatgctgg
7021 ggctgcagtg cctccgctca gccacagcac tggctgcagg gcaaaccctc aggaccagat
7081 gaggggagct cagcctgatg tccaggttcg ctggcatgtg agggacgtgg tcaggcacag
7141 ctatagccat gggtctatgg acggttcctg gatcccagga attgctggtg gggtccctga
7201 tacacacctc acgtcttcaa cctctgccct ctctgatggc ctgttggcac tcctcctgcc
7261 tatctctgca cactcagcga ctttcccctt tgatggggac aggagtggag gagggcaggg
7321 caggatggag acaccataca cattcctgct cagtgccttt gccctccctc ttccatccat
7381 gggagtcacc tcccctagac cttcantgtg gcctgcgccc ccacttctcc ctcccagagc
7441 cttcctagtg cccccaacaa gtgcctgccc tcgcctgacg ccctgtttgt ccccttctgc
7501 ctcttcttca tgctgcttgt taggttgaaa cataacatct gtgtttcaaa acacatgcag
7561 tcaggtttct gtgcatgtgt gtatgtagat atagattgtg gtgcaatctc agctcactgc
7621 aacctcctcc tcttgggttc aagcaattct cgtgcttcag cctctcgagt agctaggact
7681 acaggcatgc accaccacgc ctggctaatt tttgtttttc tttttgtatt ttttttagtg
7741 gangatgggt ttcgccantg ttggccaggc tggtctggaa ctcctggcct caagtgatca
7801 gcctgccttg gcttgccaaa gttctgggat tacaggtgtg agccaccgca cccagacgtt
7861 agtttacttt ttaattccct gtctacccac ttcttgtaag ttacatgaaa gcagacacct
7921 tgtgtgtctg atctgttact gtctctccag ggcctaggat aagcaggtgc cagtaagcct
7981 ttgttggaag aaaggaagga gagaaaatgg acctcctctg gatgggatcc ttggctcctg
8041 cgtgctgatg tgcagcctgt gacccgggct ctgcactgct cctctggctc caggcaatca
8101 tcatggaaag ggcggcctcc ccagagggct gcctgggaat gaactgggca tctggggaag
8161 cccattncct ttggctgcct cttgcctgtg gcgcccaggg ctgccgtgcc acctcctttg
8221 tgaagctggt ccccagcccc gcttatcagc cccatgtccc attcccttgg gcccagcaga
8281 ggtaaaagtg gctaagtggc ttaactgggc caggtgcagt ggctcacgcc ggtaatccca
8341 gcactttggg aggccaaggc gggtggatca cctgaggtcg ggagttcgag accagcctga
8401 ccaacatgga gaaaccccgt ctctactaaa aatacaaaat tagccaggcg tggtggtgca
8461 ttcctgtaat cccagctact tgggaggctg aggcaggagg atcacttgaa ccacggaggc
8521 ggcggttgca gtgagctgtc attgcactcc agcctggaca aggagagcga aactccgtct
8581 taaaaaaaaa aaaaaggtgg cttaactgtc ctctctgtca tggcctctcc ctccctcttg
8641 ccccagtggt atttggctct tttgtctgtc ttgtttttcc cagactgacc agccttgggg
8701 ttaaccaggg gcagggtgac aactcacagg caatgacgtc cctcagcctc cttcccccca
8761 cactccagag gctgcaggtg actcagggtc ttcgtcggtg gctccggaag ccttgcaggt
8821 gccctgccac atggccgctg tgctgctggc tgctctgacg ggtcctgccg gggctactgc
8881 tttcctgtcc cctcccccga agctgggaga tggatttcct ccctgagtga ttaattggtc
8941 tattttaagc tcatcaaagt agcctcgtac tttcaagttc ctcagggagc aggcccggaa
9001 gatgatacct gcggacttaa ggctgctcca ctgcaggata tggggtcggg ggctgtgtct
9061 gtcaattgtg aggtccctca gggcaggccc cctcctgagc tggacagggc ttggctcctg
9121 agtgctgatt ctgcaggctg tggctcagtt ctgcactatt actctggctc cggcaaatcc
9181 tcatgaaaat aagcctggga gccccgaggg ttgcaggtta gccccctaca gccttccctt
9241 gcatcatcct tgctcctcta accagggaca ctcctgcctc ctccctgact atctctgccg

FIGURE 3 (continued)

9301 ggttaacacc tgagcatctt ctggctcgat aattctgtca tgacccaata cgccttccct
9361 gaccctcaca tctgggttga gggccttcct atgccctgta cctaatcctg gttttccctc
9421 tgtaaggtgg tcagagatgg gcttgggcag agtagtttca aaggctctcc ccaaatctct
9481 gtgcagctcc cctgactgaa ggcctgtgag aacagggtgg tgcctgtctt gttcactgtt
9541 ccccaatacc tcacctagtg cctggcacag agaaggtaca aaaataaaac tttcccaaga
9601 atgaatggtg tgcatttgaa tcttatagtg tctgaggact ggggaaccat ctcttatccc
9661 tttgtatcct ctgctgccca aggtcagggc ctctgcctga agtccctact gaatgagggt
9721 taactggttc agagatgagg gaggaagcca ttgctgggag gggctagaan ctgggcccct
9781 gcccttccca cacccgagca agagcagtag gattgtcaga tgaaacacag gacatctgct
9841 cacatgtgag gaatacatct ttttttaag acagattctc tctctgtcac cccaggnctg
9901 gangtgcagt ggcatgatct tggctcattt caacctctac ctcttgggtt catgccattc
9961 tcctgtcgta gcctccccag tagctgggat tacaggcatc caccaccatg cctggctaat
10021 tgttgtgttt ttagtagaga cagggttttg cctcattggc aggctggtct tggactcctg
10081 acctcaagtg atccgctcac ctcagcctcc caanagtgct gggattacag gcatgagcca
10141 tggcacccag ctggaataca tctttagtat actgcatgta atatttggga catgctcatg
10201 ctgaaaaaac taactgaatt tgaatttttt tttggttttc ttttttgag atggagtctc
10261 actccattgc ccaggctgga gtgcagtggc acaatcttgg ctcactgcaa cctccgcctc
10321 ctgggttcaa gcaattctcc tgtctcagcc tcccaagtag ctgggattac aggtgcctgc
10381 caccttgccc agctaatttt tgtatttta gtagagacgg ggtttcacct tgttggtcag
10441 gctggtctcg aactcttgac ctcaggtgat ccacccacct cggcatccca aagtgctggg
10501 attacaggtg tgagccaccg cacccggcct gaatttgaat ttaactggga agccttttt
10561 tatttttcc tcttctttct gaatctggta accctagcaa gggagaagtg actagccagg
10621 tgccaggtgg gtctacggcc tgagtctcca gctctctatt cacattctgc ctgcttcctt
10681 gatgcttcct cctccttaaa tgctttctgt cagagaccag ggggcactct aggcctggct
10741 gcccagcacg cggaatccag gagcactcct acctcctaat ccctgcccca gctgccctct
10801 ggatggtcag ggaaggcctg aggaggtgat gtttgagctg agatttgaga gaatgtgacc
10861 agagatgatc caggaaagaa cattccatgg agaggtggga gtgaacttgg tgcatctgaa
10921 aacagaaagg ctagcgtggt tctgcgatgt gggaagggag tggaaacatg gaaacacagt
10981 cagcatttcc ttttttttt ttttttttt ttgagacgga gttttgctct tgttgtccag
11041 gctggagtgt aatggcgcgg tcttggctca ccacaacctc tgcctcctgg gttcaagtga
11101 ttctcttgcc tcagcctccg gagtagctgg gattacaggc atgtgccacc ccgcccggct
11161 aattttgtat ttttagtaga gatggggttt ctccacgttg gtcaggctgg tctctagctc
11221 ccgatctcag gtgatccgcc cgcctcggcc tcccaaagtg ctgggattac aggcgttagc
11281 acnaactcag gaggccctgg aactcatctt gtgccactca agaaaggnca ggaggggtcc
11341 tctcctttcc caccttcccc aagctgaagg actgcacact tctcaggctt ccaaggcagg
11401 agcttcggca gcaggaacag ggttgcaagg tagttagtct gggtgtctgt cgtcctgggt
11461 tgggacatct gggagctggg gcaaggcctc tgggccctag tgccagagag cagtttggtg
11521 gcgttgcctg ttctttaaaa tgggttaggt gtagccatga gccccccccat gcagcgggtg
11581 tgcaggcctc tccctcagtc agaacaccct cacccacccc tccccacttg aagcaggccc
11641 ttgagaacac tgctggcagg gaggcggctg ctcttcctcc taagggaaaa gcctgcagcc
11701 ccgtgctgct gggccgtggg gtgaggggga agccaaaggg cccctatcac ccctcagcag
11761 gagtctctcc tggctttgga aggggctgaa ggcccgggng ctctcacgag ggaactcagc
11821 cctgctcagc cggtcaggac tcctcctcca cgtctaaatc cacagaccaa acaaaggaca
11881 tggctcagtt ccaccttcaa tccacaggga ttctgcctgt tagcacgcta aggaaactgt
11941 ttcttacttt aaattatcca ttaatttcct ccaactgaaa gatactctac aaatttttctg
12001 ttgttggagg gtttaagctt ggtccctctg agggggtggg ccctgctgtg tcggcccctt
12061 gtgtctgctc gtggaaggtg tgccatgtgc caccgtgtgc catatgcagc tgacacgcct
12121 gtgcgtctcg cctggcgctg gagactgctc tctgtgtact aagcgagtac tggatcccca
12181 gggcctatct tgcccactgc ctgtggcaat nacacaggca atacacaaag cgcattcaag
12241 taagcaggtc tccctgcccc ccggcctccc agaacctcgg tgcctggagg gcaggntggg
12301 ggnagggcgt tcctgaatgg aggccggttt ccctctgcgg ggaggaaacc ccatttccct
12361 cctaccccta ggcgcttgct acatctggcc cggcatcaga caggaaagaa ccctttcttc

FIGURE 3 (continued)

12421 ccagcttgta ggaagtctga gctgggcccc ttatctaccc acagcacttc ctggaccggg
12481 ggcctggacc gggctccgaa gggnctggnc tcccggtctt ggtcttggcc tcggcgccct
12541 cgtttctctg cgctttgggc aggggaagct gcggacggca agctctcggc tttcgtgaga
12601 gcctggtgga atcggtgttc cccgaaggct actctgcggg ggcgggggct agntccgggc
12661 cttcccggca gtatcgtcnt cccccgtggg ggctggggna nggtngcccc cagggctnnc
12721 ccggagcggc ggcgtggcca aggcccgaac cgggtctgac atctagtggc ctcctgggcc
12781 cgggcagggc gagggncggg gcagggaaga aggtggaggg cagaatggga ggcgtggagc
12841 gagaaaatca ggaagcgcgg gaccaagccg gggaagggcg gcgggtctcg cccctggcac
12901 ccgctctcct cggggcccgt cccatccccc aggcctggcc gaccccaggt ccttccgtgc
12961 gcagtcgggg ctcgcaagga cgaatcccgc ggccctcgag gcaggcccgg gggagctccc
13021 ggccctccac ccccgcaggc cgcagatccc acgcaccccc gatcatgggg gccccggagg
13081 gaggtcgcga ggccgggctc acggtggccg cggttcgccc acgtgcgggc ccggagctgg
13141 cagggcccgg tcccgagcgt ggccgcaacc gcggggacct ggcacgtgcg gggcagcccc
13201 gggggccgcg tacgccactt ccggtcccgg agacaccgcc cagcccgccg cccggttgcc
13261 atggcgacgc cgtcgcgcca cggcccgcag aaccggccaa gcgacccggg cgcggcgcgg
13321 ggaggctgaa gggacgctcg ggtaggcaag gtaggaggcc gggctggggg tgggagcgga
13381 gcgcgcaggg gtgcggggcg ggggcggccc aggtgagccc tgactgcgca gggagggaca
13441 gcgcggggct cccgagtagc agccggcctc gcacctgccc cttgcggccg cgcactggac
13501 tgcggcgccg acccgcaccc tgggcccgag ggctgcaggc tcgcccgccc tctcggagcc
13561 gagcctctcc cggtccaggc ggcccctgcc ctggcctgcc cgggtgccgg gtctctgctg
13621 aagttagaac cgagaccccc gcttgctggt gaccccgagt ttggatcttt gctcctgcgc
13681 cgcgttctag agcaaggtag ggtgcaaatg gaccaaccag cctttgccac ggctgccctc
13741 ttgaaatctg aaaccggagc caccgcctct tgccccgagc agggcccggg ctgcaacgtg
13801 gagccgcagg tccccgcctg tgtctcccga cgcccccagc ttctgagcgc gagggtggga
13861 gtttcccgag tgggaaagcc ccatggcttc ggtggcctcg gtggccctgt ggtgggtcag
13921 gccggtgcca actgcgctga gggcggactg ccgccacctt gacacctggg agatggcagg
13981 gccaccctcc ctccctccct ccctgccctg tccccgactg tatcacggag cgaggatcat
14041 ccgtgtggat ctggggtccc cttcccagac ttctcctttc tggtcgtcct cctctcattc
14101 attccacctt ggcagccttt gtggattttc tgactccgta aatgaagtta catctagggc
14161 gcctagattc tgccccttg gtttaatagc taatgaaaaa aaaaaaaaaa aagctttgcc
14221 acctcctcta ggtgcctgga tttctcccgg tgaaatttga tggaggcagg aaaatacgag
14281 gcggtggcag ccccacctcc ttgcatgagg gagcctccta gagctggagg tggaaccacc
14341 ctgcttgtcc cactcctttt ctttcttgct ttattgcttt cttgctttcc ttactttcct
14401 tccttctttt ttttttgaga tggagtcttg ctctgttgcc aggaggctgg agtgcagagg
14461 gcaatatcag ctcactgcaa cctccatctt ccgggttcaa gcgattctcc tgttccagcc
14521 tccctagtag ctgggattac aggtgtgtgc cacaacacgc agctaatttt tgcattttta
14581 ggagagacgg ggtttcacca tgttggccag gctggtctcg aactcctaac ctcaagtgat
14641 ctgcccgcct tggcatccca aagtgctagg attacagacg tgagccaccg tgcccagcct
14701 tgtcctacta cttttctgcc ctcaccactt ttccttggca ggcagagctt gggctcctgc
14761 taatactggg atggttcagt ggcattagat ggcattggtg gaggggaatg agaaattaac
14821 cccctctgaa agcgtagctt cactgggaaa aagcactcta ctgtctgaac aattcataaa
14881 tgaattttgg gaatatacaa tgtcacagat taaaagcaga ccatcctgtg ttcttccagc
14941 cccaggagct tctgcttggt gtgcaggggc taatggtgga cagcaatgga gtgtagattt
15001 attagagacg gctgccatct ggctctctgt ggccccaaac ctcatgtgtc agcagtgtga
15061 ctctttctaa ggtgctctta ggaccagatt tcgtttttgt aagattggtc tttgatggaa
15121 tgcctacttt gtggcaggca ctgtgctggg cgctggcaca cctttatcca cttactcctt
15181 ctgtgttcaa taaggtagct ccagatatcc tccactttat agattaggaa gcaggcttgg
15241 agaagggact tggcttgtcc aggagctgat ggttataaat aatgagtctg ttttgaaatc
15301 acatctctct aacctgggct gcgctgggcc tttcatgggc cgttggccct ctggccttca
15361 cgggccccgt cttccataaa aaaaataata acaattggcc cggcgcagtg gtgaaagcgc
15421 gtctctacaa aagatacaaa aaattagcca ggcatcgtgg agcacgcctg taatcccaga
15481 ctcggggagt ctgaggcagg agaattgctt ggacccggaa ggcagaggtt gcggtgaccc

FIGURE 3 (continued)

15541 gagatcttgc cattgcactc cagcctcaaa aaaaaaaaaa ttaataaaaa ttatatttta
15601 caactgcatt ggtataaata caagtatagt ccagactgga ttacaattaa ttttaaaaac
15661 aaaacatttt agggctggct catgtctgtc tgtaatccca gagctttggg aggctggggc
15721 aggaggactg cttgagacca ggagtttgag actagcctgc acaacatagt gagaccctat
15781 ctctccaaaa aaagaaaaaa ttaactgggc atttcagcat agctgtagtt ccagctactt
15841 gagaggctga ggtgggagga tccctcgagc ccaggagttc gaggctgctg tgagctatga
15901 tcatgccatc gtattccagc ctgggtgaca gaggtagaca ctgttaacaa caacaacaaa
15961 aaaattaatt aaaataagtt ttaaaaaatt gcaacatttt cctcaaccct aaatgttcat
16021 ttttttcttc tgattttcat ggaaatggaa acattttggt ttgtaagcat tgggagccct
16081 ctgtcttgtg ggagaaagga gctcagtgtt tgacttcagg ctctgctgct tcctgggtac
16141 ctagttcctt ggtacctggt ccaagcagcc tggatacccc gggtccctgg gctgcctggg
16201 ccaggacagc cgccctagga taaatggaaa tgcagcccct gcggcttgtt caccctcctg
16261 taatgccttc tctcttaccc tttggtcagg gcctggtgtc tctgcatgca ccagggctta
16321 gtgccatgct ggtacaagcg tggggaggat gcaggaggac tggcatggga tggggagcac
16381 tgcctgggaa gcccccttct cctggggcta ctggggtgct aagaaagtac cctaatagct
16441 cacaccacac tttgctttct ccttcatcac atgctaggac gctctagggc aattgagctg
16501 ttttcccctt cacccatcaa gaagagatat cggcctttcc gcagtcacag tctttcctaa
16561 ctcaattcaa actctgctgt tgataaggag ggctgtagcc agcccaggct gcccgcttcc
16621 cagcctcccc tgctccgcct tccgcccccg aggctgtgca ccatgtgcag tgtctggtcc
16681 ccaataatga gattagtctt ggttgccttt taataaaacg cagtgggcac cgggagggag
16741 agcgatgctt ggctcagtga agatctgcgg gtcatgctgt ccctaatgcg ctgattgcat
16801 taagtggatt ctggctgcag gtagggtgag tgggtgggga cgagggtgac tctcacagct
16861 ctaagatcca gaaactgcca gagatctgtc accctcatcc tgagctgtca cagaggaagg
16921 caagtgactg tgtgaggggc tacgtgagct ccctctggtt gcaaggttct ggtctgcagg
16981 gcagtggagc ccttgggggt ggggagtggc agcttccagg ccttgaagct gctgcccaca
17041 gttctgctct gagaacacag agggcccaag aacagccggt gtcccagggc tgcccagtga
17101 ggagggaaga gcgagagagt attcttgctg tcaagactgg gaaatgaggg ccacgattca
17161 aagccttgct tcctagggag aaatctcacc caatgtccag gtttgcaaat gcagcagaca
17221 catttgtggg tgggtcagat tctgtccaga gataccagaa tgtctatccc tgtacacccc
17281 cacctctggc atgggacagc ccttccctgc acacaaatgg gttatcaatt atgtccaatg
17341 aatggcctca cgagagtcct gttgcggaag ggatcttccg cttatctcct gtaaagcaca
17401 gagccacgct cagaacctac ggactgtggt cgcctgcagg ccagccatgg gctttccttt
17461 catccttaca cttagatttt gaagcctcgc ctttgaatca gaagcaggct gctgtcattc
17521 cagaggccgc tgttcatctc tggcttcctc ccagatgccc aggtgcctcg ggtgcctggg
17581 ctgcctaggc cagggcaccc gccttaggag caaatggaaa tgcatcttcc tcggcttgtt
17641 tgccctcctg taaccccttc cctcttgcct tcggtcaggg cctgctgtct tgcacgcacc
17701 agtgcttagt gccatggtgg aaccgaggcg gggaggacgc aggaggactg gcacaggagg
17761 gggggcaccg cccgggagct tcccctgcgt caaagcagct tcctcagtga gctccaacac
17821 agacccttgc tggcctcagt ctctggaaag actgacagca ccaccctagt tcctgggcct
17881 tggaaaggcc aaagccacga gcaggcagcc actgttcggc aggagagcta gagatgctat
17941 tgcccacaga catcaaagca ggggtgccgt cttccgtgtg catgctggga cgggtcttgg
18001 aggcccagat tagctcctgg cttgaacttg taccaccctt acctgaaaat tgggtctgta
18061 ccaaagctca tctgcacgcc cctttagggc agggaccagg ccccctctgt gtggccctag
18121 cacctggcca agtgcaggtg gctgcgggtt gagttgagct gctatttgct gatgggaatt
18181 tgggggcctg tcgcttgtcc tctttgagcc tccatttcct ggcctgtcaa attggggcaa
18241 tgctcctgtt atccacacag tgtcatttaa gggtgacatt agagatggca tgaagtgccc
18301 agcacagagc aggtgctcac aggctgttgg gtccccaccc tactgacaca ccaagcacac
18361 gtctgccctg ccatggtggg gagcaaacaa tgttatggtt ccgactcccc gggaggccag
18421 gccgggcatt tcactgtagg atgttgaagc cgcaggcagt cgtcgctgcg agaagagagc
18481 ctggctctga tgtccactgt cttccaggga gcctggtggg aaggaaggag tgggcagcgg
18541 cccctcgctc tgcgggcctc tcctgccctt tgtactccac gaggtgtgag gaagttgccg
18601 ggtcacccag cagagggaga ggtgatgtcc cctctgttct tcctaatttg tagctaatca

FIGURE 3 (continued)

18661 gattctgcct gcagccaggc tctctgggga tgaatctaat aaaggaggcg gcgacctgga
18721 gaggaggaca gacaacaagt gaccaggggg gacagcactt ctttcagtac attttcctgt
18781 tagactcggg gtattagaga cccaaagata cccaaggagg agtgagagcg tccctgcctt
18841 cgaggagcag acagggcacc gcgagggtaa ggaggctccg ggagagcctg gggacagagt
18901 gatctgccac cgcggtgcct gaggagtgga ggcaggggtg tgggcccaca gcagggagcc
18961 aggctgggcc atgtccttgg attctgtgag cccaggcgtc tcctctagag ggcctagcag
19021 gaacagtggg ggctgcagat ggaaacgggc aggatcatgc catgcagttg gcatttcact
19081 ggcactcggg ggcacttacg caggggacag catgatgttg ggagctggct ttgtccctgg
19141 catgtctgtc acgttgaaga gggagaagct gcaggccctg acttgggtgg ggcagtgagg
19201 atgctcagga gggaagtgag gggaggggtg ccagcatgcg gcggcctggg cagccttctg
19261 tggaggagct gaggtgtcac ctgggcccgg aagaagggca gccctgctag ctagtgggga
19321 gggcatgaga gaacttggcg acgatgggag agaaggtgac ttgggacgag ctgggtttga
19381 agcgtctgtg ggatgtccag gtagagatga gtagtgaacc gtcagaaatg gggcttgaat
19441 ttggggtggc ggggacagaa ctgatacccg tgtaggtgac agataaggac agaggcagga
19501 gttccaaaaa taggatttat tgttggctgg ggcttctgga atcccgtatt cccgtccacc
19561 ctgggaatgt ggattggagc tatgtttctg ggagatgatt tgatggaatg tgtcagaagt
19621 ctaaaacaga acctagccct tgtttcagga ttccctgcag agaagtttgc agatgaccac
19681 atcctgttgt ttatttggtg gaaaactgaa atcgaaagct agctaaatgc tgaacaatgg
19741 gaactagttg gctaaaaaat ggccgtgcaa tgcaagagtg tgaggttacg aaaatgatga
19801 tgtgggcgcg tatccattga cacaaacatc catgtgtgct acgccctcag ttaaaaagca
19861 ggttagaaaa tggaataaat gatcctgctt ctataaaagg atgcatatgt gttttaggaa
19921 gaaggtgtgt gaggatatag atcaaattct gatggcctgt cattttacct tttttgctga
19981 tccatattct actgcagtgc gtgtggatta ttagcacaca cacacaccct tttttttttа
20041 attaaaaaaa attttttttg agacagagtc tcattctgtc acccaggctg gaatgcagtg
20101 gtgcgatctt ggcccactgc aaactctgcc tcctgagctc aagggattct agtgcctcag
20161 cctcccaagt agctgggatt acaggtgtga gccaccatgc ccggctaatt tttatatttt
20221 tagtagagat agggtttcac catgttgtgg ctggtcttga actcctggcc tcaggtgatc
20281 cactcgcctt ggcctcccca ggtgctggga ttataggtgt gagccatcgc acccggccca
20341 cacacacacc ccccctttaa tggtggagtt gccaacagtt gtccgatgtg gcagagagaa
20401 ctggggagca gagaaggagc ccatcactgg gagggactga aaaccagctc ctgggcactg
20461 cagggcagag cgtgtgccgg gtgtgaaggg ggtcagtgcc cctgtgctgg ccactgtccc
20521 cctagttctg atacagttgc tgccaggaga ggctggaccg tggcaagggg ccagtgggga
20581 accggggcgg gatggggggc agcagcagtg ctctggggtc tgcttgtggt ggccatggtg
20641 gtggcagttg cagtggcaac tgtgggccat ggcagcactt ttccgtggcg cacaggccgc
20701 tgccgggtga ggggagagcc ccgcttgttc tcctcctgcc gcctctcctc ccatccccag
20761 ctgctgggtg ctctccggcc ggcatgaagc tgtgcaggca ctgggcactc ccttgggctc
20821 ttagcggggc cttctcattg ctgcttcgct aggaaccacc cgctctcctt tcctcctgga
20881 gggccacgtg gggcgagcac ctgggacagc cgtgctgagc tctacccggc cccttcgcag
20941 gcctcgtttc tgggggcctt cagttttgga aacctccctg agggcccctc tagagtgcgg
21001 gtccgccccg cccctggtgt ggaggggaag tgccctgtcc ctgtgtgctg acaggtccct
21061 gtgtgcagca tggtcggggc actttctactc tgcaggcgcg gccgggcggg gagcggggtg
21121 gggggcgggc gggggcgtgg ggggagcggg gctggctctc ctgggtgtgg ggtggggagg
21181 cctcctcatc gccagcatgg agttaaaaac caggatgggg gagggagact tcccatttct
21241 cccgagatgt acttcatgac gcgtctggag aaaaggtgtg accagcccct tccttcccgt
21301 caaggagcgc gcttgcctgt gtaactcacc tgtaaagccc acccagggaa ggccagggcc
21361 tcatgaaaca gaaaacaaag ccgggcactg ggcccttgag ctcactagaa tttggggtat
21421 ctttccttgc accctcaccg catatggaac tcctggcgtc gtgtcccagg tcacgcctgt
21481 ccccgtgggg aggctgggcc cctcccccca gccactcccg agacttgaga cctctgcctc
21541 aggacgatgg gtgggaaggg gcttgcgggt aggagagcga gcgctgcttc cagcgggagc
21601 ctcgggggag ggtggcgggg ccgccgtggg aggagccgcg ccgcatctca ggcgcagtct
21661 ctaggggctg tgcgcatccg tgggggggac atgcgcatct cagggggct gctcgcatct
21721 gggggtgctg tgtgcatctc gggggggctg tgcccatcta gcggggtggc tgtgcgcatc

FIGURE 3 (continued)

```
21781 tggaggggge tgtgcgcaac ccgggggggg tgttgcgcgc atctagcagg ggcggctgtg
21841 cgcatttcgg ggggggctgt gcatatctgg ggggaccgtg cttatctccg ggggcggctg
21901 tgcgcatctt gaggggtgtg tacatctcgg ggggcctgtg cgcatcttgg ggggctgtgt
21961 gcatccgcgg gggctgtgcg catctcgggg tgctgtgcgc tgctcctctg agctctgctc
22021 tttcttgcag cgtttgcctc agcatggagg gcggggccgc ggcagccacc cccacagcac
22081 tgccttacta cgtggccttc tcccagctgc tgggcctgac cttggtggcc atgaccggcg
22141 cgtggctcgg gctgtaccga ggcggcattg cctgggagag cgacctgcag ttcaacgcgc
22201 acccccctctg catggtcata ggcctgatct tcctgcaggg aaatggtgag tcccatgggc
22261 cgctcctctt ttcccgggct tgtgggggtc cctgagaggc agtttgcagg ggtcttgtca
22321 cccctgcggt ctcttctggt tggacaaatc taagattcta gaaaagacag agacagaagc
22381 tggttcagag cctggggaga tggaatccaa acccaggctc tgaggttggt gagtggcagc
22441 tcctctccag cctggtccag cattttcacc tcgtttccac acacagcatc caaggcgggc
22501 acttcttgca gcagagggaa aggaatgagg ctccgagcca ggccctgctg caggggtggt
22561 ttgaattgag ggaaaaaaag taatcatatg tgagagtttg tatgcgtgcg tgcgtgcgcg
22621 cgcatgtgtg ttgttggtct tggacttggc agagctagag cgcctcccce tggggcagga
22681 gcaaggcagt gcttggcctc cacctgcctc caggccaggg atccaggaag cgggatctgc
22741 atccggttga cccgctcttc ctagaggtgg tcctggtgac agccattcct gagcagcaga
22801 aagctaagag gcattcctca cgtgacctgg ccttgcccac ctcttctgga cgcgggagat
22861 caggctggga tcacaaggtt ctgcttggga gccaggcgcc tgcttgctag gagtaggacc
22921 atctgcccag gcttaggtgg gggcctgtgt gaggagccac gtggctaaca ggtgaactca
22981 gaggctgctt gttgcctcag tgtgaccaac agtggacctc aaacacagcc tggaatttgc
23041 aggacacctg acaagaaggt gggtggaggg gggcgtctgc tgggccgggc agctccttta
23101 ggtggggacg agggcagagg cgctgccttc actgcctgtc ctggtggtgg ggactgtggt
23161 gcagcctccg gccctgccct ctttgtgcac agctgggggа ggttggaggg tgggggaggg
23221 gtgttaaggt gcacacagcc ctaggtgccc tcagagaagg ccaggagctg cccagggtcc
23281 ccagggaacc tggctttctc ctccggttcc tgcagctagg ctcctgttac aagccgtcag
23341 ctcacatcac ctctccacca aggtgcctcc ttgctgctgg ggcaggaggc ctcacgcaca
23401 ggcggatgca ctcacgttcc ctctctcacc tcccacagcc ctgctggttt accgtgtctt
23461 caggaacgaa gctaaacgca ccaccaaggt cctgcacggg ctgctgcaca tctttgcgct
23521 cgtcatcgcc ctggttggtg agttcccggg cgcggccttc cccgccacct gcctgcctct
23581 tcaggtatgg caaacagccg cttcacctgc tctgttccct ccccagagct gtgatgggcc
23641 cgcccccacc ccaaacatcc cctgcagggc cactaggtgc tggcagccct cgagctggga
23701 ctaaagccca gaagtcccgc tagctggttc ccctgggtca tcttacacct tccttttcca
23761 catccaaccc tgtcctcata tgacctgtct ctggcccagg cttggtggcg gtgttcgact
23821 accacaggaa gaagggctac gctgacctgt acagcctaca cagctggtgc gggatccttg
23881 tctttgtcct gtactttgtg caggtgagtc cttccaacac cccgggcctg gggccaccta
23941 ccagggaggt gggaggagga ggggagctgg gctgaaatgt acctcatgga agggcctgat
24001 ttctggggtg tgtgcaggtg agagagctgc cttcccaggc ctgtgcgggt gcaaagctag
24061 gattggcggc caagggtgcc cagctcccag gcgccccatg gctgactcag cacttttggc
24121 cacaagccca gctttctgag tgtgccaagg cctcaagcgt tgggggaccg gtcccttgtc
24181 cctaagcgct tagtgctcct cccaccactg tgctgaggag gaaggaaacc aggcagggct
24241 gaggggctca cagtcgccac gtgggcacgt gggtgtgggc ctggccgacc tcggctgctg
24301 gggtgtccgg ccgccctgtc ccggaaccgt ctcctttcaa atcctagtgg ctggtgggct
24361 tcagcttctt cctgttcccc ggagcttcat tctccctgcg gagcgctacc gcccacagca
24421 catcttcttt ggtgctacca tcttcctcct ttccgtgggc accgccctgc tgggcctgaa
24481 ggaggcactg ctgttcaacc tcgggtgagt gtcctgggtg ggagagggca gggcctgggc
24541 caccttgca cagacctcac cctgccttca gcttccccag ctgtggcttc ctgagccgcc
24601 tctcgtggcg tcacaactgg tggctgtagt tatgcttgct aagatttggg tgcttggggc
24661 ttggctttgg ttagctttct tgattttacc ctttcaaaga aacttctggg ctatgggcac
24721 cctatttatt cccaccacgc agcaggatct gcaggacaac tgcttagagc tagaatattg
24781 atctaggttt ttacattgcc catctctttt tgtctgtgag ccatagctgg agattgctgg
24841 ttgggggcgg ggggatggtt gttctcttca ggcagggcag ggaggcaggg gctggtgctg
```

FIGURE 3 (continued)

```
24901 gaatccccat ggcatcttta aggcccagga tacaaagggc atttggccta attgtgaccc
24961 cttgggcagc ttctcccttc ctctcacccc tgcaggggca agtatagcgc atttgagccc
25021 gagggtgtcc tggccaacgt gctgggcctg ctgctggcct gcttcggtgg ggcggtgctc
25081 tacatcttga cccgggccga ctggaagcgg ccttcccagg cggaagagca ggccctctcc
25141 atggacttca agacgctgac ggagggagat agccccggct cccagtgatg cgcccggccg
25201 gccctggggg ttcgcggggt gtcttcttgc ctgcccctgc tgaggcgtct tcaggactgc
25261 aggctccgga gagtggctct ggcagcaggc gggcgcgtgg gtgcagctgc atctgtttga
25321 gtgctgcttt ctggggtcag gtctccgcct cctctgcttc tcctttctcc gctgctatag
25381 accagttcat tgtgtgtggc tcccgtgtct ctgttgcccc cttcagtgca gaaggctttg
25441 ggtaggactt cgggtgttcg gtcctggtcg cagagcacag atctttaaag aagcgagaga
25501 ggaggcccca ccctcctggc agcagatgcc tggggcaagg ccaggggaaa ctgggggggc
25561 ctcagggaca ggcctggaaa ggccacgatg gctgctgaat tcaaacaagg agtccctcca
25621 gcctgaataa cacgtggcac aaatgggccc ggcctttggc agaggagcaa gtgatatgat
25681 gtgtaaagta tgttggtggt gaaagcaagg ttccccagga gaggggaggg actggccccc
25741 gggaagctct gagatgaggc tgtggcccag ctgtagtcct gaccttcctc ttctttaacc
25801 ctttagccct aggatggctt tggtgggaga ggggatagaa gcccatgact tcagacagac
25861 tttctcttgg cagatgcagg cgggcctcct cccaggctgc tccagacatg ggggttgggg
25921 atgggggggca ccttgcagcc ccttcctgct ggggctccct ccttgtagca ccccccttgc
25981 ggctcagctc tggtttcctc tcccaggctc acccaggctc tgctcaggct gggaggcaga
26041 gggcacaaac cttataattt tttaaatgaa aaaccgctgc tgctggctgt ggctagagcc
26101 ccctggggct gctggagctg ctgcctctgt tctggaggac gagccttctc cttatctgct
26161 gcccatcttt ccaggaagtc aggatggagt cagaacaact acagtcatcc cccgtggtgt
26221 ctgcacatca ctccagcccc ataaagagtg tcatgttagc tgagtcacca tttggcttcg
26281 gcctggaaat agtgtgatta gaacactgat cgtgtgcgag gccaggagat caagaccatc
26341 ctgactaaca aacacagtga aaccccgtct ctactaaaaa tacaaaaaaa ttagccaggc
26401 gtggtggtgg gcgcctgtag tcccagctac ttgggaggct gaggcaggag aatggtgtga
26461 acccgggaga tggcgcttgc agtgagctga gattgcactc cagcctgggc gacaggctca
26521 aaaaagaaaa aaaaaagaac accgatcatg tgcttcttgg atctggtgac tgttctctcc
26581 ctgttttcct ttctttttgg gtgtttgagg agcatggctt agcttgagac acacacagac
26641 actgtgtact tcagtgaagg gcttaatata cagtttccaa acctgacgac tcttcttctt
26701 gtaatggctg ccctttttcct acctgaggcc gtcttagaga aaggggccag tctcctctaa
26761 tgctcagatt tcccatagtt ggcttttgct gtgtctcctg cctcaggcag tgtcatttct
26821 gggagcaggt ggttgtagtc caggcccctc cccagcaggg tctgcccagg ctccttcgag
26881 ccccttttccc cgcctcctct cagcctgtcc ggatgacagt gttcgcctcc tgtttagact
26941 gtacactctt caggggtagg ggtgccgtca gttcttcaat cagctggcac acacttgtat
27001 agtgaaatgt ttacatgtgg gaaaactccg ccttagacaa actaccaaag tacaatcgtg
27061 tctctctcta gccggaatgc tacagagaga aatggaacct tagatttgca acaaaagtct
27121 gtaaactggt ctgtttgcca aagtgaacac tggatgacta aggagctgaa gaaggccccc
27181 agaagcggat ttgtggtggg ttattttatt ttgcctgtgg ccaatcttct gtgaaataca
27241 atgtgctgtt ggtgcaacag atgattcaat aaatgtctac agcagacctc tcgcctgtta
27301 tcttccttta ctgtggtaat aaaaggagcc ggagctttta gcagccagac acatgacgtt
27361 agctctaggc ctgagagaaa ttgaatatcc gcaggctggg cacagaaggt aaagcgatta
27421 gtgatattga tcatggccta ggggctgaaa aggcccagcg gttgtccagt ctcgacccag
27481 cagaggcagt gttgtttcca catggttaga taagcccttt cctctcagcc tgagagggtg
27541 gcctggatgg tggagctgca agagcctgat aagagccttg gcaaggaagg tcccccagtg
27601 tttaatggac ccctttccct ttgaaatcag tctttttgat ccttaagaag aggagcaaag
27661 cttttggaac gagctgagat tccactttag atccacgtac gtggctcagg cagaaggtga
27721 gtttttggca aatttgggtg ggatcgtact gtgcgctttg tgccttgtca cttaaagcag
27781 tggccgccaa cctttttggc agcaggaaca ggttttgtgg aagacaattt ttccacagac
27841 cagggctggg ggtgggggat agttttggga tgattcaagc acgttacatt tattatgcat
27901 ttttatatta ttacattgta acatataatg aaataattat acaacttacc ataatgtaga
27961 atcaatggga gccctgagct tgttttcctg aaactagaca gtctcatctg ggggtgatgg
```

FIGURE 3 (continued)

28021 gagacagtga cagatcattc gattctccta aggagcatgc aacctagatc cctcgcatgc
28081 gtagttcata attaataggg tctgcactcc tatgagaatc taatgctgca gctgatctga
28141 caggaggcgg agctcaggtg gtcatgctca cccgctgctg ctcacctcct actgtgcagc
28201 caggttctaa cagccacaga cccatactgg tctgtggccc tggggctggg gacctctgac
28261 ttacttaaag caatttaaaa actcaccaga gctcactatt taaagggacc caacagcaaa
28321 tcctctaata aggtaacagg agtccctgct tccccgcgtt ggtctcaact ttgtcttctg
28381 cattgtagct ggtctctaac gtgtaggagc cagaacggag aggtctttcc ttagggcacc
28441 tgtagagtcc ctcctgaatc aaggcttcca tgtggacctt tatttttatt ttatttttt
28501 gagacggagt ttcgctcttg ttgcccaggc tggagtgcaa tggcgcaatc tcagctcact
28561 gcaacctccg cctcctgggt tcaagcgatt ctcctgcctc agcctcctga gtagctggga
28621 ttacaggcat gcgccaccac actcagctaa ttttgtattt ttagtagaga tgatggggtt
28681 tctccatgtt ggtcaggctg gtcttgaact ccctacctca ggtgatccac ccgccttggc
28741 ctcccaaagt gctgggatta caggtgtgag ccaccacgcc cagctgagtg tggaccttta
28801 tttactgaaa agcattcaaa gcacaaggct tgatacatgt cacaaactgg ttcctcagtc
28861 agcatctcat ttctgttttt ttttttttt ttttttttt tgcaccagcc tctgaatctt
28921 atgggtctct gtgaatattg aagtgatccc agtcaccaag ggatgatgcc tccctctcca
28981 gccagccact tcacctcttc ctctacgaat ctctggctgg ataatagcag ggacctcatt
29041 ttcttactgg ggtcccaaag cgactcctct tgactgtggc atagagttat ttctccaaac
29101 aagcacgcgg catttcccaa gccatcctcc tggacccagc gcctgcactc cgtggggcgc
29161 cattcgggtt gcttgctctt ttcagaccat cccttctgag agaggaggca gcgtggtgga
29221 gggaagagct gtcagtgtgg ctgattcaga ggcctacctg tgcctctgtg accctgagca
29281 acttcagctc tcaggatctg cttctgcatt tgttttggag acagcagtgt cgtctgtcga
29341 ggtgtgagga atacattagg gaacacatgt gaagtcccta tcacagtgtc tgggacacaa
29401 tggacctccc aagaggatgg cctgttagaa tggactctga gcaccttctc ttctatgtgg
29461 gccctctctt catcacaggg ccttccagac atgatacctg ttatcagtgc cccatcttat
29521 tcctgagaat ggaataactc agctgccaag actccaagct cctccaactt gtttttaaag
29581 ctgaaagagg gtgactccat tccccgctgg ttcgcctacc cattcccagc cgaccgcagg
29641 ggagtttggc caccatgtgc acgtgcgtac tgtgtctgcc tgcagttacc aaaaggtacc
29701 catgagcaga tatccagggg aagaaagtct gacttactta gtgcttcgga ggatctacac
29761 agacttctgc tgagtttgct gtttgcttct actctgagta gctttttatt aaggttgagg
29821 gtaaattctt tttaagaagt attaatattt taaatgaatc actccttttc ctttcctcca
29881 atgtcccagg acttcttaaa aagctggggg tttggccagg tgtggtggct catgcctgta
29941 atcccagcac tttaagaggc caaggcggga ggacagcttg aggccaggag ttcaagacca
30001 gcctgggcaa cataacaaga ctccatctct gcaaaacaaa acaaaagtta gcagggcgtg
30061 gtggcatgca ctgtggtccc aggtacttgg gaggctgagg tgggtgggtc acttgagctc
30121 aggaggttga ggctgcagtg agctatgatc acaccactgc actccagcct gggtgacaga
30181 gtgagactcc atctaaaaac aaaacaaaac atctaggtcg ggcacagtgc ttatgcctgt
30241 aattccagca ctttgggagg ctgnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
30301 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
30361 nnnataaaaa taaaataaaa taaaaataat tattatgcac aaagtagtag agtaagagca
30421 gtaatgagtc caggaagtgt tgttaccatc tataaccctt cttcatctga cacaaagtag
30481 agatcgctga aacctagcag aatgaaaacc atccctaaag acaatgaatt agaaagttct
30541 ggccaggcac ggtgcctcat gcctgtaacc ccagcacttt ggaaggccaa ggtgagtgga
30601 tcacttgagg tcaggagttc gagaccaacc tggccaacgt ggtgaaacct catctctact
30661 aaaaatacaa aaattagctg ggcatggtgg cagacacctg taatcccagc tacttaggag
30721 gctgaggcag gagaatccct tgaacccggg agtgggaggt tgcattgagc caagatgagc
30781 caagatcatg ccactgcact ccagcatgga tgacagagtg agactctgtc tcaaaaaaaa
30841 ggtcttttct ggccagacat ggtggcccac acctgtaatc ccggcacttt gggaggccga
30901 gtcaggcaga tcacttgagc tcaggagttt gagaccagcc tgggcgacat ggtgaaaacc
30961 tgaatctaca aaaaatacga aaaaaattac ctgggtgtgg tggtgcacac cagtggtccc
31021 agctacttgg gaggctgagg caggaggatc gcttgaaccc tggaggttga ggctacagtg
31081 agctaaaatc atgccactgc actccagcct gggcgacaaa gtgagacttt gtctcaaaaa

FIGURE 3 (continued)

31141 aaaaaaaagt acttttccaa aatacttatc caggccaggg gcagtggctc acacttgtaa
31201 tcccagcact ttcggaggcc aaggtgtgtg gatcacctga ggtcaggagt tcgagaccag
31261 cctggccaac atggcgaaac ctggtctcta ctaaaaatac aaaaattagc cgggtgtggt
31321 ggtgtgcacc tgtaatccca gctactcagg aggctgaggc aggagaatca cttgaaccca
31381 ggaggtggat gttgcagtga gccaagatcg tgccactgca ctccagccag caacagagtg
31441 agactccatc tcaaaaaaca aaccaaaaaa acccgaaaat acttctccgt atcccactga
31501 aacactagtt aaggtattct aaggtaactg agagctaaat ggtgatttgc cactcctttg
31561 gcttggtggg gctggaaatc cctttattc tcccatgttg taatgagatg tctttgctca
31621 ctctttttt tctttttctt taaatctcag aaggaggccc caaccctgag cacaacagca
31681 acctggccaa tatcttagag gtgtgtcgca gcaaacatat gcccaagtca acgattgaga
31741 cagcactgaa aatggaggtg tgtactgttt gacatgcttt ttattgatca cagcctctac
31801 cgggctcatg tctggatggc caacaaacac actgtaaatt atcagagggg ccaagtgcag
31861 tggctcacga ctataatccc agcactttgg gaggctaagg tgggtggatc acttgaggcc
31921 aggagtttgg agaccagcct ggccagcatg gtgaaactcc atctctacta aaaatacaaa
31981 aattagctgg gcgtggtggc atatgcttgt aatcccagct actggggagg ctgaggtgga
32041 ggatcgcttg aacccaggag gtagagattg cggtgagccg agattgcacc actgcactcc
32101 agcctgggcg acagagcgag actctgtctc aaaaaaaaaa aaaaaaatta tcagagagaa
32161 cggtagcaat gtagttgacg ccctcagcct agagtcttct gagagctcac cgtccttggt
32221 cattggttct aggagcacaa ctgggagtag gatgcatagg cagtatgcag ggccatacag
32281 tcaagaccgg gactgctcag ccaagttgta gggagatgag gtattataga aagaaggggc
32341 agtgtgggag aaagccctta cgttaggtgg aggaagtgtt tgtttttcag atgtttatgc
32401 cctaaaggag aaaagccagt agaaagggag aggatggggg tgatagagat ggtgggaggg
32461 gcaggtatat agagtccacg taagggtaca cctttctctg gggcaggtag ggaggctctg
32521 aggaagtaat agttatctta atagctcaaa attattgagt gcttcctgca ggctaagcat
32581 ggataaaaac tctatgtgtg ttatctcatc gaatcctcaa aaccaccgta agtagaaact
32641 gaggcacagg aatgttaggt gagttgactc aagtcgcata gaaaatgaca gacttgaaac
32701 cagatagtca gactttagag cctgtgctct aaccgttatg cttattgcct tcatgtaggc
32761 catggtgcag agaaggtggt ggagacagca actattttcc ccacctagtg gccctttta
32821 cctttgtgaa agaggaaggg agaagttatg agttgagaat tcctggaaag ttgcaggaca
32881 gggtgcttgg agtggggaat attgaataca tggacaaggg atgagtaggg gcatggaagg
32941 tccagaagag attagtccct agagctctgg attaattgtt tcaattttga gaccatgctg
33001 ttttcagcag ctcttacgac aaggcctagg tgtgtgatat ttcacagcag tactcagcag
33061 cccagggtag ggaccttgag acttgaccgg tgggtttctt acaggatggg ggtgtaggag
33121 gatgagtctg agtgggattg ggtagggact agatgtggca tttggttagg gatggtagga
33181 agcaaagctg agagatggct gatggattgt tcagggaccg gagaaacaga ctccttccaa
33241 gttaggggta gtacttcctg ttacctcaaa agttggttag agggtgggga caaaaataaa
33301 gattgagagc tcaggctcca gcctagagag cccatcccat ggagttagtt ttgcgatact
33361 tgtagttctg ggctgacatg tgggaaggaa ttgggaaaca ggaaaatgtg cttgtgttgt
33421 ttattgcaga aatccaagga cacttatttg ctgtatgagg gtcgaggccc tggtggctct
33481 tctctgctca tcgaggcatt atctaacagt agccacaagt gccaagcaga cattagacat
33541 atcctgaata agaatgggta agtgtgcgtc tgggaggagt ggtaggggac agagccttta
33601 tgttccaatt ctctgcaagg caagtactgt tgatctctgc tagtgtttca gggttttgtt
33661 ttttgttttt ttgttttttg agacggagtc tcactcttgt cgcccaggct ggagtgcaat
33721 ggcatgatct tggctcactg caacctctgc tttctgggtt caagtgattc tcctgcctca
33781 gcctcctgag tagctgggat taaagatacc cacagccatg cccagctaat ttttgtattt
33841 ttagtagagg tgacagggtt tcatcatgtt ggccagactg gtctcgaact cctgatctca
33901 ggtgatccac ctgcctcagc ctcccaaagt gctgggagcc accatgcctg gccatgtttc
33961 aggttttaag cacacttgct ccttagcaga atctaatgag ttaattgact ttcccttaat
34021 gtagtttcta ctagcaggat cccaaagact tgttcatcgg atgtagaagg ggatctttgt
34081 cctattatcc ccatctgtag ccatataatg cctaagtctt aaaatgctcc accaggactg
34141 gcatcttaat ggcacaggaa attatcagct aagtctgctt cctcccaact gcaatcaggg
34201 cagaaaatcg agcgtgggga cattcttggg ctgttccctt acccagcagc agctgcattt

FIGURE 3 (continued)

```
34261 tctccttgag gatgcgatct gctccatgtc tgtgtgtact tcatagactg ggtaagaaag
34321 agacagtgat agaatgatga gctttatgga ggctgagctt ctagaggcat ccactcatgc
34381 cagcctgttt ccttccctgt cagaggagtg atggctgtag gagctcgtca ctcttttgac
34441 aaaaaggggg tgattgtggt tgaagtggag gacagagaga agaaggctgt gaacctagag
34501 cgtgccctgg agatggcaat cgaagcagga gctgaggatg tcaaggaaac tgaagatgaa
34561 gaagaaagga acgttttaa agtaagcatg aaaacatggg tgtttggtgg gcttccggga
34621 ggaccctgat agatgcctta tgcatgcctc ttggtttctt ccttcatgtg ccccagggta
34681 taggtgagac agttcacata ttgtacaaac atttattaag tgaatactga ataggaccng
34741 actctagtga ggactatgag aaagaaagag tcaaaancat tgaaaacgat gtccctgcca
34801 gtgaagagct caaacccagt gatccattcc agtactgaca ttcaggactt tgcaaggtcc
34861 tgtgtggctt tgtctcatta cctcctggct cctcacctcc tgactttcct ttatccctag
34921 tttatttgtg atgcctcttc actgcaccaa gtgaggaaga agctggactc cctgggcctg
34981 tgttctgtgt cctgtgcact agagttcatc cccaactcaa aggtgcagct ggctgagccc
35041 gacctggaac aggccgcaca tctcattcag gctctcagca accacgagga tgtgattcac
35101 gtctatgata acattgaata accaggctac atgtgccccc gggttccttc ctagaaatgt
35161 ggcagcccat tccagcacac aggcttctgc agcaatctct gagggtaaag ccggtgggag
35221 gctcagcnag ccaggaggcc caaggacagg acttgcgacc ttgaagccaa aggaatctca
35281 cttgtggggc ctccttgtca gctctgctgc tgtctcagag ccatctggat gagtgtcccg
35341 acaccctctc ggatgcaggg caggaccacc cagctggtca gactctgatg ttgggtagct
35401 ggcctctgtg gggattgtaa gtgccctgag gcgctctgta ctagaaactg ctcttaataa
35461 taacggtgat tattggttgc tgcattgctg ttgtatggct cttgagtctt cctgagtttg
35521 tgtccagctg ttgggatcct ctggactaac tttcaagtcc ctaggcttag ctctactact
35581 ccaatcccag gattattttt ttaattggcc atagaaatgt agttgtatcc aggcactcca
35641 tggtctgatg ctggtcttac agcctaaacc tgtcattcat tcatgcaact aatatttctg
35701 gaacactagc tatataccaa gaactgtact tatctagagc ataagctatt caactggtgt
35761 cctgtgttag atcctgctct agaaggtagg acctgaacac aggactgctc tttactggag
35821 ctgtcccagg gcagtgtgag ctgctgtgca agcatctgtt gctctctttg tgtgacttct
35881 gtgcccttct ttcaccaaac agcttaaggt tctagcccag taggaaatca agtgccagtg
35941 ataagggact tggcagtggc tcacacctat aatcccatct actcgagagg ctgaggcggg
36001 aggatcactt gagcccagga gtctaaggct acagtgagct atgattgcac cattgcattc
36061 cagcttgggt gacagagtga gaactcttct ttttttttt ttttttttt ttttgaggn
36121 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
36181 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnng ctgtgtgttg cctgcagttt
36241 gtctgcttcc ctaggctttg ggaaatgcag gtatgcacag caggaatgtt tgtttagcac
36301 aaatctgatg tcaagaggaa gtggtctccc tcctcgcttg tccctgaagt aaacacagag
36361 gtagcagtta aacaggaaag gctgcaggct cagtgctgct cccccgcttc cccactccct
36421 tttagatggt attgagtctc cagccagttg caggaaacct gccacttttc cattcacaat
36481 gcagctgcct agctctctat atggtgagga ctggagttcg accagatttc tgatgaaaag
36541 acaaagcccc tggaagcagt gggcctccag ttgtgacgaa gcccttaggt ctgcagtggg
36601 agctgtgccc tctgctcgct accctctcag catgcagcct ctaaacacag ccactcaagc
36661 tggcagccca cctcagcctt ggctttggta gggtggctgg attgggtgca gcttcagcta
36721 caaaatgaaa caagagctca gtgctccctg ggtatagcta gagagaaacc tcttccaatg
36781 aagaaaaagc ccccctttct caacaaatgt cacaggcctg tgctgacagc tccttagaac
36841 ggcgtctctt gtccacagct ctcttgaggg cctaccacag tgtccaactc tgaggagagg
36901 aggtgctcaa taacgtgacc gagcaaataa atgcacaaag gggtggctct gaatgtccca
36961 caagagacca agcaaatcat cgggaatttg gccttagccc taacctgccc aattcctcct
37021 gagggagggt gcccccatca ctgcatcctc aggagggtaa gcttaaaagc cagctccctg
37081 aagaggctgt aatggaagga gaaaagcatc acaacaccat ggttttccaa gtgttagcca
37141 tttataaata agtacatttg ctttcataca tacagttcct tgtacagatg acaatctgta
37201 tacatggggc aggaaaatgc attcatttga acttttcaca tctatctcac acagctcaca
37261 tgtacagaca ataaaactgc tcaagcaagt acagcaaagg aaaatgtctt tccttataca
37321 caggggtaga tgcctctgtg gggtgtgggg catcccactg cacggcttca caactgtgtg
```

FIGURE 3 (continued)

```
37381 gtgttcaata tatcaggaga gagaacaaac atgcattgga taatatactg tacagagaaa
37441 gtcctttaca tctgagtcat agaaaaccta aaggaaaact aagtgcatta aagctttttc
37501 cagcaagtgt cttgaaagga cagcaaagag gaggaagaat caaaatcata ttagtacaaa
37561 tcactcttta attgtagact gtacatgtct gtactaatta aaatcatctt ggatttggag
37621 gagacagaac agagacaaag atgctgtgct agatggaaag gaggccacgc ctgaaaaggc
37681 acctgccctg agcctgatga ggaactggcc tcactcagca ggaatcagcc aaaggaaaca
37741 aaaaacaaaa caaaaccaga acaggaagtg taacttacag gatttccaaa atcacctgtg
37801 aatgaagtgg agatctggag ccggcatctc ttaacttttt ttttcccccc agtaaattgg
37861 tatgcaataa ggcaggtaca ttcaagtact gaattttcca gaattaactc ttgtctggcg
37921 ctggggacca aagggattga gttgagcccc ctctaaccag actttctggt tagcgattag
37981 caaaagaaaa attcagccag caagtgctac aaaaacaaag cagctagggc acttccgttc
38041 cacagagtag gtctacctgg aaaaatgagc gcggcgctgg cctgatctct acgcgtccag
38101 cggcagcctg gcaagtcagc tcagcgtcgg tatcagagtc agcaggaggc aatgagatga
38161 tggggtgagg aaacatgaaa gtaacacttg atttttggtg tccaattatg cgttcatttg
38221 gtactgactt tcaaagctct gactgtggcc accatgtggc cacaagcatc tcagggtggc
38281 tcacctgctg tgaggctttg gacaccgaaa tgaaggttac caacacttca gcccttgagt
38341 ggtctgtatg acaaaaggag ttgatgaaaa cccagtgatt attcaagtag ctctgcacag
38401 tggctccacc agccccattg tgcttgtgtc cagcccccag ccagccacct tttcttggga
38461 gcagccagag ctgagttcaa ggcattgcat ggtgaagggt tccatgacac gtctttgcag
38521 gtagctcttg cccctaagcc ctttgttcat tgttgttagt catggtagat ggtcggtctg
38581 gaattcctag aggaagagga gaaagagctg cacctcccag tgagcgagca gcaggaccgc
38641 acagccctcg gtgtggaagc agcagccggc tgcctttgca ggtcggtttc ttgagaaatg
38701 ggcagcccag gaacagacag gcagggtgtg cgaggctgct gggcactggg agcatcaaga
38761 ggaggctggg cacaggggcg ggcacctctt gccccctggg atagcctatt ccattttgtg
38821 gcaaagattc agtgagcact ggttttgtcc aaggcatttc tgcagtagaa aaataactct
38881 ctctgaatca aaccacccaa ctgtgacgct tctggagtta taaaagctgg aggctgagag
38941 gaactgacag gagggaggct catggggaag aaggggcttc tccatgtacc catccctcta
39001 ccagagcaag ggagctctgg tcaaccttct tccagcctct gcctggctga agtccaaagc
39061 atgtagtgtt caaagagttc gtcttgcaca actggcacag atgcacgaag accccttcca
39121 ggcctctgtt gccttctcct gaacctctga gcccgcttgc ctgttgccct ggaagcagct
39181 cctctctccc gaggggatgc cgctgttctt tgctacaaaa caagcgcagt gcagagccac
39241 agtgacacct agtggtaaac tatggtgaag cacaagtgac atccacatag cccactgtac
39301 gtgactaaaa tctaaggaaa aatacttatg gatattaaat tagatactga ttaatttta
39361 atttttttcta ttgggtacat ctctggaata taaaaatacc aatatttaga gaggggctca
39421 taaactacta tacaatatta aggactgaga ataccttttct ctaagctgtt ctgtttacaa
39481 taatttagga aaagtgttta ataatccagg cttaactaca ttagcaaact tgtatatctg
39541 gacaacgacc tggggtactg tacatgattc taattagtgg aattttcctg agccatcgag
39601 tttaagttat acacatctaa aaagaggggg cacatggggg aggagcgaga agggtttgtg
39661 ctctattaac ttgggacttt aatagtgcac gtctgcaacc cggacaagat acaacagcag
39721 atacaaaatg gtctccattt tgtcatgcca aattctcatg ttacacaggt tttcccttta
39781 ctttgtaaat aaacattaat tgttaattgt taattgtgtg ccttactgat ccagtagata
39841 aagtaaccgt gtctcaggag tccctaagaa cattgctgga aaagcacttt aaaatcactg
39901 caaatatttt tcatattaaa aaattcttaa tctttttgat gcttatatac aagttatttc
39961 ttgtgctata aatgttgtga tccactgctt gatgtctttc ctttcctttt ttcttgaaaa
40021 atacactaaa agacaagagc ggttctgcta ttttctaatg aagacattac tcacacttaa
40081 atatccagta cttcagttac aaattcaaac agtaaagtgc acccatttat agacatgatg
40141 tgatagaaac ccattagtgc aagaatcctg ggccaatgga acatacaact tggtgagaaa
40201 cctattaaac tgaagtttgt cacctctgtc ctccnnnnnn nnnnnnnnnn nnnnnnnnnn
40261 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
40321 nnnnnnnnnn nnnncagggt gaaagataat gtgttctgga gtcagttctg gtttcaaagc
40381 atgactttgc ttcttgctag atgaatgatg cacaacgaca ccacctctct tgccttcagt
40441 ttcataattt gtaaaaggca gataataata tccactttgc aaagttaaga gctattatta
```

FIGURE 3 (continued)

40501 taatatgaaa cactgtaaga gtactcatca gagagcacct cctgtagtag taaatgcttc
40561 acattttcag ttcagtcccc tactccctca gggtttgctt tgggactaaa aaccactgcc
40621 acataggtaa gcatggaaag aataagctaa acaccaagtg cagagaaaaa gactcaaggg
40681 tcaaagtgga ccacaggctg accagaagca tgccaagagg ctactgtgaa agtaacctgg
40741 atcagcctgg gaactggatt tgggcacctc caagaagagg tgggnagana acagactgca
40801 ctgaagggtg tgccaagtgc tgctgagagg gtaggtcagc accttgtaag gaaaggataa
40861 gggggattcg gannenggaa ggctgaagcc tgaggtagaa ccccagtgct ggcttcctat
40921 ctngggacag tttggcttac tggcatatgt gtgtgcatgt gtgtgtgtgt gtgtgttagt
40981 gggaggggct aaacattgtc ttttctttct ttctttttt ttttttgag atggagtttc
41041 accttgttgc ccaggctgga gtgcaatggt gcagtctcgg ctcactgcaa catccacctc
41101 ccacattcaa gtgattctcc tgcctcagcc tctcaagtag ctgggattac aagtgcctgc
41161 catcacgccc ggctaatttt tgtatttttg gtagagacgg ggtttcacta tgttggccag
41221 gctggtttcg aactcctgcc ctcaggtgat ccgcctgcct cgacttccca aagtgctggg
41281 attacatgca tgagccactg cgcccagcct aagcagtgtc tctgaacaca gaaggctgca
41341 caagaggaaa tgggctctga ttaatgtagg aaagactgca gtctgacatg aggaagaaat
41401 tccgctagaa tataagctta atgaagctct gctcaaggtg tctagtgtag tcaggtgtgg
41461 agagggagcc tggagcagga cttccttaaa gggagatggg ggagcccaga acaaatgaag
41521 gagagggcac tcaccccagg agcttcccgg caatcttgga gttgtagatg tcacactggt
41581 gcagtggacc catgtggccc gaggctttgc acagtgtttc gtggaactgg aactggagca
41641 ccagactgag gaagtacctg gttgagaaac gtgagggcag gctggggagt ggagaccagg
41701 agtggcaaac tcagagccac ttgggaccag gcagctctga catgcagtgg ccacgagggc
41761 cccatgtcag cagtcactgc ctgagcggag gctccttggg cagggccagt gcttgggctg
41821 accctagagt gctgggggge tctccttctc tagctttctg aaccagcagt caccccaacc
41881 cttgcccccc tctgacctgc tctcctccaa gggacagcag ggatggtgta gaatattctc
41941 atcagctctg ccctctacac tggacctgat gcaaaagctc aaaggattct aggtcagagg
42001 cctcagaagg ctgcgcagcc caacaggaag agatgggcac gggtacaggg tgacatttgg
42061 acctagggcc ttcctgcaaa tcctcacccc agttcccatg cccctcaact cctaggactg
42121 gcactgctaa cagggcccag acctttaccg tatgtagggc acaccagcag aaaagtggaa
42181 cttggcacct ggatcaaagt cttcctctga gtgaggaata gcggggcaca agccctggta
42241 tttcaacctg gcagggaaaa ggcagagggc ttggtggtgg tgtctggcag gttctcagag
42301 cctctttctt ttccttcctt ccttccttcc ttccttcctt ccttccttcc ttccttccct
42361 ccctccctcc ctccctccct ccctccctcc tttcttcctt tccttctttt cttctttctt
42421 ttttttttt tttcagagtc ttgccctgtt gcccaggctg aagttcagtg gcgtgatctc
42481 agctcactgt aacctcccca tcctgggttc aagtgcgtct cttgtcttcag cttcccaagt
42541 agctgggatt ataggcgtga gccacncgca cctggcctgc ctttttcttt tcaaagcacc
42601 tacgtgtcta tcaaaactgg acttggcagg ctgaggatac ctaatgcatt tttattgatt
42661 aaataaatga aaacacatgg gattgggaat cagagtaaaa caatagcagt taacactgac
42721 tgaacacttc ctctatgctc agaaccatgt ctggctctcc acatagactc acgcatcatg
42781 cactcaacac cgtcctggtc tctgggggag ggaccagtgg gggagcgggg gaggcaaggc
42841 ccctgcttgc atcctaacag tgggggacag attatagacc aataaacaag gtcatttcat
42901 ctagtaagaa gtcttgagaa gaaaaaagag taacatcact gtggggcagt actactatta
42961 ttcccatttt gcagatgagg aaaccaaggt ttggagatca gggaattggc ctcaggacac
43021 agagccagtc acagtcacac cctgataacc tgaaggtctg tttttgcccc tctgagatgt
43081 gcgactgtgg gaaagttgct gattgtctcc gagtcttgtt gtaaaatgtc tataaaatgg
43141 aaggaagtag gccaggtgtg gtggctcaca cctgtaatcc cagcactttg agaggctgat
43201 gtgggaggat tgcttgagcc caggagttcg agaccggcct gggcaacaaa gtgagaccct
43261 gtccccacaa aaaatacaaa aatgagccag gtgttctggt gcgcatctgt agtcccagct
43321 actcaggtgc tgaggcagga ggatcgcttg agcccaggag gtcagggctg cagtgagcca
43381 tgattgcacc actgcactcc agcctgggca acagagtgag accctgtctc aaaaacaagg
43441 cttcatcatt ggctcactgg agcctctgaa ccatctccaa ggtcaacaga gtgtcaggag
43501 cttccctccc agttgacagg gccttattca aggcccaggg agttgccacc cctcctgctt
43561 agccccaggg ttccatcagc tggggtgatg ggcagggctg acagaggcct cctgcaggtg

FIGURE 3 (continued)

43621 cctcagatgg cctatgaggg tgaccttaga gaggcacgag gtcaagactc tgccaagccc
43681 accagggaga cctgctcctg aagtcaaggg tgggagtgag aggacaaggg gaggagctcc
43741 cagctctgtc acctgctcgt ttgctgtgtg acctcgggct ggtttgtgac ctctgtgagc
43801 ttcagctgtt gcttctgtga gctgagcgct ttgaagtgat cacaacgcaa ggtggttgta
43861 aatgttcaaa ttaaacataa gaggccgggc acagtggctc atgcctgtaa tccccgcatt
43921 ttgggaggcc taggcaggca gatcccttga gcccaggagt ttaaacctgg gcaacatggc
43981 gaaccccatc tcttaaaaaa aaattacaca aaactagcca gtatagtggc gcacacctat
44041 agtcccagct actcaggagg ctgaggaagg aggattggtt gaggctggga ggtcagggct
44101 gcagtgagcc gtgatcatgc cactatactc cagcctgagt gacagagcga ggccctgtct
44161 caaaaaacaa acaaaaaaag taagagcctg ttgcagagcc aggagcaccc cctggcaacc
44221 ccgaatcaca ctatgaccag cagactgaag ctggtcagtc cgcggcctgc cacatatgga
44281 ggagccctgg tcacaggcat ttattatccc atcaaatggc caggggtcag tgatgaggtg
44341 taacacgggg gtccctcact cgagagtgat gcctcccaac tgctggtggg ttcccatggc
44401 agcttgggga tgttaagccc gaggcaggca tgcgtgcatt ttaggggcct gcctctctcc
44461 catggggagt ctttccccag aaaggaaagt gctcaggccc ccgcacactc cattgcctgg
44521 caccaggaag ggagggaggt ttgggagggc ctgcctggtg gtcagtgttc tgtgattaga
44581 gaagctcagt tgcaatcatt gtgtgagctt cccagctacc cagggagcag gtggggccat
44641 caccagggtc catttttcac agatgaggag actgaagcct agaaaggcta aaagacctgc
44701 tcagtgcctg tgtggaagat tagagaggtt cttgtgctgc agaaagtcct ggaaatggcc
44761 catctggtgg aagatggtgc ccaacccacc tgaggttcca ccactcctga ttgtagatgt
44821 ccttccagat ggtgccgtca aagaccttcc agcgaaacag gtccatcagg tagccaaagg
44881 ggatgaaggc gatcttctcc agggcaatat gcatcaggaa attgacctca tcctctggga
44941 agagggtgag ggcagtcaag aggcaacccc tgggatctcc ttcccgtagc tttgcccgag
45001 agccctgacc ccaccaagcc ctagctgcgg tcccaggctt gtgctgccct cctgctgggc
45061 ccttgatcct catcacctga gtcctggtgc tagaggctga gcaggcctat gttgagcagg
45121 tgcttgtggg aggaggccga gagggtgatc acagacccca cagcctcttc aaaggctggg
45181 ttggcacctg tgcggaagat gatggagagg ttcttgtact gcaggaagta ctggaatggc
45241 ccatttcgtg gaagatggag agcgggtctt ctgtggtcac ttccgcgcac ttctttattc
45301 tacaagcaaa gggcatgggt ncaaggagca tggagaaagg gagtttcccc attcagactc
45361 cttgagggat tgggtggtcg ggttggagcc cagacaggcc ttggcagttg ttctttttt
45421 tgtttttctt ttgagacgga gtcttgctgt gtctcccagc tggagtgcaa tggtgcgatc
45481 ttggctcanc tgcagcctcc ggctcccagg ttccagtgat tctcctgcct cagcctccta
45541 gctgggatta caggcacatg tcaccatgcc tggctaattt ttgtattttt agtagagaca
45601 gggtttcacc atgttggcca ggctggtctc gaactcctga cctcacgtga tccactgncc
45661 ttggcctccc aaagtgctag gattacaggc gtgancaccc atgncccggc cggcaattgt
45721 tctttcttac tccaggcccc atatccaccc agcaaccaag ccctttcctc tcacgtcacc
45781 tttccttctg tgtttctcac ttgccttctc cctcctctgc cctggctttc ccctgcccat
45841 acctaagaca gccagacagg gcagaggaga ggaccgggtg tgtggtgccg ctgtgagcac
45901 ctgaaatcgt cgtcctggta gaagttccag gcagagatgt gacactccac ctctcgccca
45961 tcggttggcc tcattagcat caacttttc cagaaactgg gtggggcagg tggcagcgcc
46021 aaggnaggtg aagaatgtct cagcctcttc caacattttc tctggcttcc agtgctatgg
46081 gagaagaggg ggtgtggggg gcgctgccct gtttgtccaa ggatgccaat ggggaggggc
46141 agggacagac aggccaagca ctaacctgga ctttcatgat ctttgtgaca tcctctggga
46201 tcttcttcag gaagggcagg accgggtcta agatgttgac ccaggactga gccaacgtgt
46261 tctctggaag aggaggagag gggctaagag gaggctggag acagtctctg cccttatctg
46321 gcactcacct cggccaccag cagggccttt acccaggagg tgggcaggga tgggccccct
46381 caggtcgatg agctcgggcc catagtggcg gtggagggcc ctgcgcacgt aggtgtgcgg
46441 gttcaggtag agtggccgca gctcctggaa tagccgctcc aggtcttgct ccagggtatc
46501 cgactcatac ttggagtgcc acaaggcccc catgtctttg tnaacctagg acaggagaga
46561 ggactcacca ggagctcacc atctcaccct ttagccatgg cctagctgac aggataccca
46621 gacgccttct agggaagcca ctcaacctcg ctgaccgtct gggtcttctt tggcaaaacg
46681 gggagaatac ctgcccattt cagaaaggca ctgagcaaag tccatgatcc gagctcccca

FIGURE 3 (continued)

```
46741 ccatgcctgc aaggccctgn agtcccctct tatcccacct catccttagc cctgtggcca
46801 cctgggcttc tttcaggtct tccagagcac cggctcctcc tgccgcggag cctcntgcct
46861 aggctgctct tctttcacat cttgcttccc gtcttttgcc catctttcag gtcttggata
46921 aatgtcagag tnccttccct gtccacttgt ctgcagcagc ctcttctcat tctctttctc
46981 agcagcctct tcctattccg catcacccca gatggtagtg agagacaggt gtttactgaa
47041 tatggtctgt ctccctgcag gaccgtaggt gcatngangg gcaggtgcct gtcttttcac
47101 ctatgtgtcg tcagaacctc cccagcacct agcacctagg tgnctctatg aatatttgga
47161 gagcccagct gcagggctca taggtcctgt ctccnagggc cctacggcca ccttcactct
47221 gggaaagctc agctttgccc cttccttgcc atcactccaa gctctgtggc atcccatctg
47281 ctcaccgttg agctgtgcag ccttgttgct gagctccaca tagtgctcga aggtggtgca
47341 gatctggcgg cccacngcat cctgccagcc ctgccaggcc cacagcagct cctctttgtc
47401 cctggaggtg gccatgactt cgagttctat gcccagaaaa ggagccatgg gggacccact
47461 cccacttgcc tcttccctca ccctcctggc aataagggag ggtggttggg ggctgggagg
47521 ggcccctgga agggactgag tgctggagat tctgtggttg ggtgggacag ggcttgggtg
47581 ttcaccagac tccagggaca ggcagggccc ctcattcagg cacacctggg ccatactgta
47641 tgtcatctcc aggntaggcc agaagctcgt tatactgggg ggacaggtgt tacccagggt
47701 aggctggtgc cccaggtccc cccccagtgc tcgccccacc agccaggaga gatcggtacc
47761 cccacctcta ctcctctcct tctccagagg aagttctgat cctgcccttc ctccactccc
47821 tgcctccctc ctgccccagg cactctgtct cccctaccag acccctagga ccagggagca
47881 tgtgncctaa aggcctgggg gttccatcca tcacctcccg cagctcgtcc ttggacagag
47941 ccgccttgtc tatgttctgc agcttactca gcatgccatt cacatccggg tccttgaact
48001 gggtgacttt aaacaggtgg gcctgggtgc caaagtaaat catgaactgg gacctctcca
48061 tgtccttgtg cagctgcaaa gagagccacc aaggttgggg gacaggggaa gggtcccact
48121 tgacctccag ggccatggca ccctaactga ctccctaatc ccattggctt agggagcaag
48181 atttgctgag agggagaggg cagaaacctg gatctctaaa agcaaggagg tgagatttac
48241 aaaggacaag aagcaagaga ttcttgcctg gaggaagatt agggcttcca ggaggcccag
48301 agcttgcagc cctctttgac atgtacactg gggcttccct tgttcctctt gcttgggatg
48361 ggggcaggag gtgtggtgcc ccatggtccc aaatccctaa aaagagcaaa gagaggagag
48421 gctggggtgg aggtggtatc acatcatctc ctcctgattt ttcctggtga tattggtgac
48481 atagttccaa gtggcctcca tgaacttgtt caacacaact tcacctgttt ggtcataaaa
48541 ctgcaggaag atcttggtct cggtctcatt gtagaagtca tctgcaggga aggatatggg
48601 catcatgtgc tgagaaggtc ctgtccgcac ccagtccctt tgtctggtcc catgcttctt
48661 ggggtcccag ccacacctaa actgtgttca cccttgatcc tgagccaagg caagagctgc
48721 ccataacaga ggagcacaag tagggaaggt ccagggcaag tccaccttgc tcccatgacc
48781 aacagagcag cagagctggg tgaggccagc tccatattgt gacgtcagag gccagaggct
48841 cagctacaga atgttagaac tggaaatccc catccacctc ctggtccaac tgtccccatt
48901 ttacaggtgg ggagactgag gtccagaaag tcaagccaaa tgatctggtt catgtagtta
48961 gctaatggcg agccaaaaag aggaaccagg ggtcttgact tttaggccag aacttgttct
49021 gctctcctcg gccacttttc cctggggctt gagccaccct tgggttatgg ccagggattg
49081 atgaggccct ccctccatag gctgggacca cttctgagca ggtggatggg atgatggatg
49141 gatggatgga tggatggatg gatggatgga tggatgtgat atatctgggt ctctggaagc
49201 caggttatct ctctaggttc ttctaacttg acctcttatg ttcacatgtt tgggcttttt
49261 ctctcttatg ggctgtgcaa tcaaccccag ggctgtgcag ggccctcccc atcataccta
49321 tttctccagg gagaccctga tgttatgagt ttgcaattag gagagggcac acaaggtagg
49381 gactgggtgg ggtgtggaga ggtgaggcgg gcaccctcca tnccagggtg catgcccatg
49441 agtacctgag cctgatctga aaaccagctg gcaagacacc agctctctgt aggcctgtgg
49501 tcacccccag tactcttgca ctgaggtctg tggaggccag ctccatctgt ctcaccttgg
49561 atttctcctc agactcactt ccctccacac tcagaatccc tgtaccctgg gttgccttcc
49621 atcactctca gcccacaatg gctcacatct tcaaaaaggc cctacctcta cacatgggct
49681 gataccacga ctagagaaac tgaccaaata acccccccaat atgatccatt atgggtaaaa
49741 tatatttata acatgaggaa agaaagaaaa aaatagtgac acttctggtt aagctgatcg
49801 actgagctca caggaaagtc tcccttcctg cttcaagcat gtagacagtc cattctcaca
```

FIGURE 3 (continued)

49861 ttgctataga gaactacctg agactgggta atttataaag aaaagaggtt tagttggctc
49921 tacagttcca caggctgtac aggaagcacg gctggggaag cctcaggaaa cttacaatca
49981 tggtggaagg cgaaagggaa gcaggcacat cttcacatgg tggcaggaga gagagagaag
50041 ggggaggcgc tacacacttt taaacaacca gatcctgaga actcagtcac tatcacgaga
50101 acggcaaagg agaagtctgc ctccatgatt caacacgccc acctctcctc taacacaggg
50161 gaggtcaggg tgcatccatc cagagagctg cagcattggt gtgggggat agctgatccc
50221 gatggaaatg gcgtctctgc caagctggga gcccctgcaa gcgtcaggtg catgcatggg
50281 attgtgcagg ccattccaac ccaactgaag atggctgata ggtacagttt accaagtata
50341 agagaaaatc cagttttagg agggacaata ataattacca accctcactg tgggttggtt
50401 ccaggtgcca gggaccatct gggggctttc aacatatgaa tcctcacaag aaacccatga
50461 ggtaagtact gtgatcaacg ctcccatttt acagatgagg aaagagaggt acagaacggt
50521 tgaggaattt gctcatagtc acacagccag cagatggtgg aagtgggact caaattcata
50581 tggttggttc cagagcttgt gctaaactgc agacacaatg aatgaagcag actgcaatgg
50641 aggaaggaca agtagcagag caatccaaaa gggacttaaa atgcaggtct ttatgccatg
50701 tataaaaatt agctcaaaaa tggatcacag cttattttct tttttcaatt taaaagagta
50761 catttaaggc tgggcacaat ggctcatgtc tgtaatccca gcactttggg aggctgaggc
50821 aggtggatca cttgaggtca ggagttcaag accagcctga ccaacatggt gaaatcctgt
50881 ctctactaaa aatacaaaaa ttagccaggc atggtggcac atgcctctaa tcccagctac
50941 tcgggaggct gaggctggag aatcgcttca cctgggaggc agagtgcagt gagccagatc
51001 acaccactac actcnagcct gggtgacaga atgaaactct gtctcaaaaa ataaataaat
51061 aaaataaaaa taataaaata gtacatttaa caccttatct tggacaacgt gacagtcatg
51121 agatctttgt gtnggagaag aanggggaa caaaagagca aagtctacaa aattcttaga
51181 acacagaggg aagtttcatg acattggatt tggcaatgat tgcttggata tgacatcaaa
51241 agcacaggca acaaaagaaa aaataagcaa attgaacctc atcaaaattt taaacttttn
51301 catatnaaaa aacactgtaa cagagcaaaa agganacaac ntggaatggg agaaaatatc
51361 tctgactctc agatcnacga cgaatnaanc gccagaaagc aaaaagaaat nnnnnnnnnn
51421 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
51481 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gaaagaatta cccccgctca ttcttgtcta
51541 aagccccccc cttgcacctc ttctctcccc cagtcccgga gctggtaacg gccctggccc
51601 catgagcagt ttgccttctt gagtcactgc ctgtgtagta catacctgac cgggagtcca
51661 aaccaccttg gtgctctgaa gtccactgac tcatcacacc tttcttagcc tggctcctct
51721 caagggcatt ctgggcttgt aaacagacat aggaagcctc tgtttaccct gaagcaccac
51781 tgtccagccc attggttccc actggcagca tggtagagct gagagaaaca ggctctcagg
51841 gtacctgact tgaggggaat cgtttcatga agctgaactt caagcatatt tccagtacat
51901 tctttcagag tctgtttttc catccaaata taagccccag gccattccac ttagtgtctt
51961 ttcaatgata ggcaagaatg atatctgagt tgaacttcgg tgcttctgtt gtttgagttt
52021 actgtgcctg gtggtatatt gggcattctt tggattgagt gttctgaggt gagagagtct
52081 tcccgaggca tcctgtctgt gcttccaacc ctgaacaaga ccttacatga gagatggact
52141 gatggactgc ggcaatcctg ggctgtcaag tggatagata gttaaaaagc attatactgt
52201 gggtaatgaa aagggagagg aaaaaaaaag aaggaaaagg aattatagac ccccagggtc
52261 agnccagtta agagctctac ccacacctgt caacccctct ctcccccagt ttaggttctg
52321 agcagtattg gacttgtagc ctgcagttgt cttttgactt gcaggccgca ggtgtctttc
52381 tgttatgtga atgagttcca tggaggggca tatgtgtgat tccaccgtta gatgagccct
52441 tggggcagng cagtttggga tgtgctcttg ggggaaagtt ggctgtttcc ttgcgctctg
52501 ctcctacccg aaggttttta agtccctctg aattgnctnc atctgagatt agtagagtag
52561 caggcctgaa ggatgatggt tttgtcctct ttggttctca cctgcttgag aagtaaaaca
52621 gtaactttgt tcttctgggc ccttaagctt ttttggttaa gtcttccttt tcagaagtag
52681 atgtcattat atgccaaaag tctagctctt tgctttacca tacagggacc tgtcccaaag
52741 aaaaaggctc ttttttttagc cagcatattt ccccttctac cctttttacctt tgttgttctg
52801 attttaggac tctggctggc catgtgcttg tggttgcctc tcctgcattt gccactggat
52861 ttgcactgca tcgtttggag atacaaagcg agcagttctt ggtcagaacc ctcctctgct
52921 tttcattgtg tttgataatg gttactgggt ccttctctca agggtagcaa ggccaagctg

FIGURE 3 (continued)

52981 atggctgctt gtttaggagg ccatcagttc cttcctgtgg agaagggtct gaaatggaag
53041 tcagtggtag aaggggctgg tctgctgggc agggcttaca tccactgagt tctaagattc
53101 ctttcctgat ctgcacctac gcctggtctg tatggtggaa tttgtcagct ggaactcaga
53161 aacaacaact tgaaaaaaaa ataataatta gaacatattt gcataagata gctatttact
53221 ctggaaacca acaacttttg agatttccct tgccctgtgg acgcccagct cctgtcatcc
53281 ttccttaggt cctgcagtac agtcttcccc tgaatgccac cggggaccca gggggactcc
53341 acccccctaa gcaagcacac acatactcac agttgatgag ttgctggtct ttgagtccca
53401 gctctcttac cctcccttta ctccaccagc ccgacgaccc atgactgagg aggggatttc
53461 tacagtctca ggatttagaa agtctgtaag ccatccatgc tccagaaagc accgatctgt
53521 tgtagttgca aaaacaactc tgtaatttgt tgaggttctc aaactgacag ccagcgagac
53581 tgggtgggag gccctggatc tgttctccct gactgcggga ggagcagcca ctaggacttt
53641 agcaggaagc ccacatggag gctccgcagg ctgtggccca gctggtgatg gccctttgc
53701 tcctggcagc ctgaggcaca gctgcctgta ttgtcctcat ctgttctgac tgaaggatgg
53761 aggtgctgaa taaattaggc ctcaggcctc taccaccaga gagctggaga atgggtccac
53821 gtcattcaag gacctgaatt ttttatgctc aggagcattg gaatcctctt cttccaggga
53881 ggaattagcc tgcaaggtta ggacttgaag agggaaggta tttaataact gggcgaggat
53941 gggtgtggtg gctcacacct gtaatcccag cattttgngg aggctgaggt ggccagatcc
54001 caaggtcaga agatcgagac catcctggct aacatggtga aaccccatct ctactaaaaa
54061 tacaaaaaaa aattagccgg gggtggtggc gggtacctgt agtcctagct actcgggagg
54121 ctgaggcagg agaatggcgt gaacctggga ggtggagctt gcagtgagcc aagatcnggc
54181 cactgcactc cagcctgggc gacaggagca agctccgtct caaaaaataa aaaaaaaaaa
54241 aaaaataggt gaaaattcct tataaatcca ggattggctc tgagagaact ggctaagatt
54301 caggaagaaa caaaaaattc agaatcctac aaggttttga tgacaattag ggccaaaatt
54361 ttaggaggag atgtaggatg caggagaaaa ttaaagtgtt ttctttatat cagaggagga
54421 aatagtagag gtcagtgaag gtctggggta gggaaacatt cagactgtcc attgcatggc
54481 tgtggagtga gactgccctt agcctgggcc agccttcctg cgccacaaat tgggcatccg
54541 tgatgctagg taactgtggg aacaaaatga cagcttagag cagccatggg tgatgtttgg
54601 tggtaaaaaa cctacaggcg tttggggtcc catgattgtt ccagaccatg actcttcctg
54661 gttgtgggtt tgttacagag caggagaagc agaggttatg acagttatgc agactttccc
54721 cctccttttt ctcttttctc ttccccttgc ttttccactg tttcttcctg ctgccacctg
54781 ggccttgaat tcctgggctg tgaagacatg tagcagctgc agggtttacc acacgtggga
54841 gggcagccca gtactgtccc tctgccttcc ccactttgag aatatggcag ccccttttcat
54901 tcctggcttg gggtagggga gaccattgaa gtagaagcct caaagcagac ttttcccttt
54961 actgtgtgta ctccaggacg aagaaggaag atcatgcttg atacttagat tggttttccc
55021 agggaagagg gcggagcaga gcaaagtcac tgtgaaccct gggccaggcc ctggctgggc
55081 cagctcctga gagcgtctcg tgttgcagac ccttgcccac ttcacccacc tgcaccttct
55141 ccccctctca cagtgtcact gctgctaatg gtcaaagtca aatgtgtggc cacatgggat
55201 gggccaggtc ctctcaggct actttctgga tgtcattttt aaaatatgga aacatgcagg
55261 tgccttccca aagaggcttg gactggtata tccaacgaga aacaaataag ctaaagaaag
55321 tttaaactca agaagaaaga tgttgacagt ctatgtaaca gctggaaagt ttataggcac
55381 ccacctttgg gacaacccag tgattatgaa catgtgatat ctactattta aaagaaatgt
55441 tctcaccttg ggttgattgt ggtataccat gtgttatgaa aattgttgag ctgaagcttt
55501 gaatcgattt agttgagtct gactcacttg ctttggttcc tgtgtatttt actacccctc
55561 ttgtcagtga ccttccttcc ccaccccacc cagagtgaat ttgtagcatg attgtataaa
55621 cctctatgta gaaaatggag atttcttgct ctgaaatgtt aagctctaac tgatccattt
55681 ctgtgtcctt tagcctagta tgtctgaact tccattcttg ttatatattt aaactttccc
55741 tctatattat aggttttgtg gcatccacgg tcaggtgtag aggaagctgc cccttgcaga
55801 actgtactgt aatatttttc ttttataaat attttcacag gactgattgt acacagggct
55861 tgtaataaaa ttttaacact gtgctgtgaa acaactatgg ggaatctcca ttgaaggcta
55921 cttcatgggc acctgaaagt ggagtgttat agctatgact ttctatttct tgtttcctaa
55981 gtaaattaaa cctaattttc acccttтcat tctgtttcag cctcctgtat aagaagtacc
56041 gtattttctg cccatcatac tttgtaataa aacttgaaca tgtatagatt gactgaattt

FIGURE 3 (continued)

```
56101 ctttttgtga cgaattcagt tcttcccatt ttgtcacttt ggtcatgttc catagacagt
56161 ggcagtgccc cagcccgagc cggncagtnc tactcatcac cttttatatc ctatttagtc
56221 attcacttgt taattcagtg gagcacctcc tctatgccag gggacatggc cctcaagttg
56281 gttgcagtct accagaggag acatctacat aaatacttgt ggccgggcac agtggtctgt
56341 gatcccagct gaggcatggg ggtggctgag gtgggaggat cacttgaggc caggagttca
56401 agaccagcct tggcaacata gcaagaccct gtctcaatca atcagtcatt tttattaaaa
56461 aaaaaaaaat ttcttttagc cagactaaga tctaagacaa atctttttt ttgcatttct
56521 tttttctttt tttgagacgg ggtctcgctc tgtcacccag gctggagtgc aatggcgcga
56581 tctcggctca ctgcaacctc cacctcccgg ttcaagctat tctcctgcct cagcctcctg
56641 agtagctgag attacaggca tgtgccacaa cacccgactg atttttgtat tgttagtaaa
56701 gacggggttt caccgtgttg gtcaggttgg tctcgaactc ctgaccttgt gatacgcctg
56761 cctcagcttc ccaaagtgtt gggatggtac cgagctcgaa ttcgtaatcg tggtcatagc
56821 tgtttcctgt gtgaaattgt tatccgctca caattccaca ccacatacga accggaagca
56881 taaagtgtng agcctggggt gcctaatgag tgagctaact cacattaatt gcgttgcgct
56941 cactgnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
57001 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnntattc tagtcaaaaa
57061 caactggcat gcagcagtga tacctactcc ctcttatgat ggccaaaaag aatggcattt
57121 tgactgttga cattttaggt gttttgccca tagggtatag gcccnaaatc ttnttgncct
57181 ggaaaactga gagaccctgg ggagaagcgg catcagcggc atatttgtaa agctacagga
57241 cctgctcatt cgcctggaca ggtgagtgcg ggtggtgtgt ggcgttccaa atgagcagga
57301 gcactttcct gacagacact gagaaattct tcaaaatgca gtggctcctc agcgagaggc
57361 ggttttaagg agtaagatgt aaagatgcca gcagtgctga gacctggcca aaaagcaatc
57421 tgagaggcca ctgggctcct ggctagaggt ccccgctaga gctacacatt tcatcttcct
57481 gatgtgagag agaggggtat agagatggag agaggacgct gatgcctacc ctggtctagc
57541 tgagaaggcg tcgctgatgc ccaccccggt ctagctgaga aagcagtcct tagctatcag
57601 ggagtgtgga aagagactgg ggcttgagtt agtcttgcta tttgctttcc agccaccaga
57661 gacaagtcct gccctgtttt aggtttgggg gtggagaggt ccgtctattt cgcagaatct
57721 aaattagtaa agaaatgctt ctgatctcag gttcatgctc cctgtgtctt aatgatccct
57781 tcaaagggca cttacttact tacttactta ccagacacat gggactggtc gcagccagat
57841 ttttattcag gaaactgccc ttcctgcagc ccttccctaa cagagcagaa tgattcctaa
57901 caacagagaa cgaatgctcc tttgtttcgt ttcagttgct tagttgagtt gtttagttgt
57961 aagagctgat gtacagccag tgccaggcct gtgccgaggc tttggttcca atgttcttgc
58021 tatccactca atagtggcct cgctcactca aaggtcagcc taccaaatga ccacgagacc
58081 tagaagatgt gtcttctaga caagcttcaa gatcagaagc ttgctgtcta tctcatgccc
58141 ccacccaggg ctgcaaaaca gggctgaaag gggtctttct tcaccctgaa acagtttttg
58201 ctgcgtctga ctttaaaata tcacactctc ataaacactc ctagaaaaag aaaagaaaca
58261 gttcctgtaa taatagataa tcaaactata gatctcaaaa tggaagtaga caacaacggg
58321 catgaccctg ggtggaaata ttgggactca gtctcatggt cccaagagat tccaatctgc
58381 ttctgaaaga aacaatttct gtatgacacc atctgggttt acaggaagag ggccactaga
58441 acaaaatctg gctgtaggtg aagatccaac taacgggatt gtcacaccag ccagccttcg
58501 gttagagttt ccttttagtg gagtatgaca ctcagtgatg tatcaggatc taaaaagtag
58561 atgtctccag aacagacttc ctatagcctt tcactgagaa tccagaacaa tctttttttgt
58621 ttttctatgt attcattccc tcccccaccc tggaaaagca gcagctttta gtggttaatt
58681 gtcggctaaa aatctagaga ataactctca gtgaattctc ttaatagcca caattctcat
58741 ctaatttcct tggctgactg caatatttaa agggctggct gtttttgaca tgttcctaca
58801 atgcctcttg gaaaatcaat gtagtgacat tttacataat tatattgggg ttttcttcca
58861 atttagcact taattcttct ctctctgatc cttttacccc ctaggaagac cagcaataat
58921 cctaatgcac actgcagacc tggtggctgc agggcgctgg cggtctgggg aggggccaca
58981 gcatgagaat gcgctgcggc tggctgtgtg ggtgacactg gctaaggatg caaaggcaaa
59041 ggggcatcgc actctttcta aagtttctag aagaaaacga acaagaacac cttatttttt
59101 aaaaaaggaa aaagacaatt acacaacaag aacatcagtg aaagcgattg tctcctggaa
59161 aaagctgacc agtgtgtctg atctccctgg gttaaagcac acagcacgaa aatctcctcg
```

FIGURE 3 (continued)

```
59221 ttggcatctg aagagaggga gagagcatgt ggctgggcac agacaagaga aactgaagtc
59281 ctggctcccg ggaggccgct tcccagtcgg ctgcatgcaa aggatttcag catgtgggct
59341 gccacctctg aacaccacac agaaactaca tgagaaatta agccagagaa ctacctcgaa
59401 tgtggaaacc aagcctgaaa atgtgcagtg aaaacatatg ggtctcactc caacaaaacg
59461 tcttttaaac attagactcc acgaacggtc tctttggttt gataggggat gtctgccctt
59521 gccgagagtc tcgggacagc tgcctgtaca ggttgtcctg gtatgcctga gccacaggta
59581 gagtccgagc taccttcacg tcgggatagg aggaggcctg gctgactcgc tccaagaggt
59641 ctccacgaga cccgttagcc agcatcccat gggggctgta ataatcgtcc tccagcaaat
59701 ggccattctg tgcattgttg gttttgttat aaaaggcaat gttgctgatg gatgatggtg
59761 gactgaagga ctcaggctga ggcaggttgc cnggagacgt gggactgagg acggtgtcca
59821 cagatgacac tgcccaggtc cgattctgct ggtggaggtg ggggtactgc tgagtccgtg
59881 ctgttttatc tatgatcccc atgaatggtg tggtcctgga tcgggtgggc tcactggggt
59941 aacctccttg agnttgngag acactggcga cagctcngtc acatgacctc tcatatgcag
60001 gtttcagtgg gatctccatt tgctggatgg aagaagatgg ccggaaggtc ggagttaggt
60061 tggaggtgga ggagatacta ttgctagatg gagagaagcg aaggcctaca ctttgggaat
60121 gcagcaacgg cctgggagtt ggcgggtgcg gtttgatttc gttagggttg acactgatag
60181 gtcttctgat taaatgcgac acatcagggg aacctagttg gctggaggag cgctccagat
60241 ctagttttga gtgacagact gggtaatagg atgctgactg gctgcgtccg atctgtggtg
60301 tctggctgta tctcacgcca cctcggtatg ctgaggaagg tcgctgtgga agatcctctt
60361 tggtcaatcc tccatgctga cagatggcgc ctgcactcag gctggcttgg acatgctgca
60421 ccggccttcc atcccctacg attcccccaa ttgacccttg ataaaccaac cggctctggc
60481 tgactcctat gtctccttgt gaagactggt atttactggc cattgaatgg gctacttggc
60541 cataggctcc tgtggggatg acggtgcttg aatggacggc tgggctgggc tggttacttc
60601 tcacaatctg ggccttggca ggctgtagcc tgagcccttg ctggggaact gccacaggga
60661 gctgaggcat ctggtagggc cgctgggcca ttttnggata atggagattt gggtctgcca
60721 ggaagctgac tgacgcttgt ttcctgctga tagcgcacgg gtgaaccaga ctgggatcta
60781 tagacactca tactttcagc tggagggctg gcccgatact gagggcctct ccggagaggg
60841 gaagggggag ggcttgggta ttctttgccc tgtcctccca caggaggggg gctgaaacgg
60901 taagatcctc cctgatgagc cggggaagtg tgtggtgggc taggcctgta atgtgagttc
60961 tgatgagttg gagaaagggc aggtgaggtg gactgatagg tctgtctagt gggagagcct
61021 acagaactac tggaccggaa gtcaaaaacc tgatgagagc caagaggtga gctggagatg
61081 taggctgagt cccgatgcgg gggagaggag ggtgggctct ggatgaccgg gagcccctgg
61141 ctgggccggg cctctgtctg gaggccanat ggaaacattt tcaancatcc tgtntccagg
61201 tactcctcct cgaatatatc ctgtacagag tatatgtctt catgctgtgg ctcaggttct
61261 gcttcttccg gcaactgctg ctgaggctgc ggtggcggcg gtggctgctg tggctgctgc
61321 atctgtctac actcttcttc cactctcagc agaagtctct ggatctcacg gttgttggga
61381 cacagcttga tggcctcgtt caggtcctct aaggctgctg cgaactgtct gcttgacaaa
61441 atgtgaacac aggaactgga gttattctga atccagagaa atcaccaaga gctgagaaca
61501 tggtcactat ctcagagagc tgtggtggcg tctgtgggag tctggggtat ttcggtttag
61561 ttggatttat agggaaggaa acaggaagac tcattctgga agagtccgaa tgattcccta
61621 acccattcgg cctccatctt ccttctggcc tctctgggaa tagtggagaa atgccccatt
61681 ttgtcagtcc angccaggcc cttcgggaag ttgttttggt gcccaaacac acaccaggca
61741 gtatgagagt tgggtggggt tgtcatttac agggtcactg tcaagtagct taggaacaaa
61801 aaaaactaca tatttaaaaa tcttgctgat ccagcagaaa tttttcttta tattctcaat
61861 attttagctc caaatttcag gcctgttcta ctctaggcaa gaaccctgtc ctctgccctt
61921 tctaaccctc atctctcaga attgctgtgt ctccatttgc tctatcaaag tggggcaggc
61981 attaggcttg ctgactaggt cactcggggc tggagctgag ctgactccaa ggagactcta
62041 atttacaaat gaagaaatga actgtcagag acattcccat cttctctact aatatcacct
62101 gacaggggca actgaccttg tgaaaaagga actnacccag aggnaangga gaatttcagt
62161 cagatttctg tgaggacttc gttgccccag ctaggtaaga gtgcgatcac tgccaccacc
62221 atcactagcc tcaagctggg tgtgattttc gggaccatgt gcccctggcc atagcagtga
62281 acctgaagca gttctttccc atgcagctgg ggaggaggag ggcaggaggc ataagagaca
```

FIGURE 3 (continued)

```
62341 gtgacaagtt acacaaagat cactgatggc aggttctagg ttattctggg agcttttctg
62401 cccagtggcc aaactaaagc agatgggcac aaggaatgtg gaatggtgtt ctcttagagg
62461 gtcctctcca tcctggcagc aaggacctct tgactctgtg tgtgtgtggt gggtgatggt
62521 ttaggtggaa gggggtctga tgccaccaga cttgaggaga gccccagagg gaagactaac
62581 taatccggga ttcctcaaac gcctatggga gaaggaagag gaacctgtga tgttaagtga
62641 accatcctac tttcccctcta aagcgggggt ggggtgctct gcttaagtga tttctttaag
62701 tcgccgttgt ctctttgtgc tagctgctct gggggttgaa aaatgcctga gatacggcag
62761 aggcttcttg gaagaagctg ccattctcca gtgaaaaagt gtttttctcc cgaagtgaaa
62821 cattcgtggc ctccacagcc cagcagattt gctgatcttg agctgagggc tgtgctatgg
62881 gagccatccc caaagcttgg ccctgacagt tctcaaggac agagatgtat ttgggtacat
62941 atgccccagg gtggccgacg ctgccttggc caaattttca gaaaagagac ctcttgcttt
63001 caccctctct ctctctcctc acctgctgct gcgttttgcc ncttgntctc gcatagtaag
63061 cttcataaga tttcggtttc agctccaggg ccttagtagc aaattcctcc gccattccaa
63121 aatcctgtca ttacaatagg tgtgcaaatg agtcattgat gagatataaa aaatagacaa
63181 cttgggaata aaacaaatta tttgtacatt ttttgacatt catggatgga acgcctggga
63241 cgctagacgt gcctgacatg gctcaagccg cgtggcctgt ggactgaaag ttccttctgc
63301 ctttgtcctg tgtagcgctg cgactgctca ctggacttta tctgcccaga atattaaaat
63361 agttttaatt ttccctcttt ttgcctttct aataaagttc tcagcccgga agccaaacac
63421 caacccttte cagtgaagct cctgaagtta tccatgcatg taggaggcaa acgcacacat
63481 gtccacattt cacgtggacc ctggtctgag ccccacacct tcctgctaca tgcactacac
63541 tgtggctttc taagggacag gtgagaaggt gtcctgtaaa gtccgggagg atgacgtctg
63601 taaaggaaac tgcactgact acagagacta gaggggagagg gggtgtcaga cagacctggg
63661 ttcaattctc tgaccagctt cataaccccc aagtgtatag tgtattttct catccatacc
63721 taccccgggg gtaggatcaa gtgtgataca gtatttaaaa gccctgcaca gttcccggca
63781 catagtaggt gctcattaat tatcgtnttt ctttctattt cacagaggct tctctgatct
63841 gttccttttt ttaactggtt ccttanctta ggaaacactt aaaattaaaa ataaaacaaa
63901 acaagagaag accccagctg tgactgaggt ctgtgctctc cctagaggaa cacgctcaca
63961 ggcatgtcat tcctgagaac ccctggaaac agagactggg gggtggggtg gggaacaacg
64021 ccgactaggc agncctatgg ttttcagtat tcccttaaat actaagctaa aagagaaggc
64081 tttccccttc ttttagtgtc aaagccacag tatagcatgg tagcctctgg ttatggtcat
64141 gagtcctaag gcagaaaaca aggccttggg cagaacaaag gccctggggc ataggtaaaa
64201 aaatttttta aaagttgatg ggaaatcctc tacctaagaa gacagacgtc tggtgtttgg
64261 cacagtattg gcccaggagg tgattaacaa ccgcttctga acaaatgtgg agctagaaag
64321 aatttcagct tttatgggtc accctctccc tgctctgctt ccagccttac cctttattt
64381 gttcttttgg gaccatcatt attcctaaaa atatgagggc atttttaaaa agtctattat
64441 ttggtaggga aaagagatca gtgggttggc agcaacttct gagctttgcc tcatcctcat
64501 cggctggtgc aggggaggcc ttggtgctgt tcctttcatt tcacaggaga ccccaagagc
64561 atgaaaatct gtctcccgag accgccacct tctcccttcc cggggttggg gggcctcttc
64621 cacatcccca cttccctcag caccctggag caactctgcc tctgagggcc atggcgttgt
64681 tctctgagac agtcaggttt aactctgcct ctgaaggtca tggctgtgtt ctctgagata
64741 ggtttaactc tgcctctgaa ggtcatggct gtgttctctg agacagtcag gtttaactct
64801 gcctctgaag gtcatggcta tgttctctga gacagtcagg tttaactcag gttccttcca
64861 gtgtcagggt ttagggtctg gaatggatga tctaaaggtt acacacacaa atctttctgg
64921 aagatccctg gccatgaggt tcaaatcttg cttctgtcca gctgagtttt taattgatca
64981 atacatctgt ttggcttgga ttatttttat ctcctgagga actgtgagac ctcatgcagg
65041 tagatgtcag agactcctga cacaatctta gctatttcca aaacctgaga gagggttctt
65101 ggcttggagt gtcttcacgt taaagggcag tggctgaggg agcagtgtct ggggacatgt
65161 gagatgggtc tgcaaggtag gcaggaagga ccgtggctta ccagggaaac aataaaggtn
65221 cagggggnat ttctgacatt ccccctagga caagaaggga catagcatgc tgggatacaa
65281 aaatgaggaa agaagcgaaa agttcgatag cctcaaaaat taaaccttca aaaccatggt
65341 ccacccccte tccttaccce ttcttcaaat gataagggaa tatacctgaa attaatgact
65401 acaattcaca tctaattgca actggaattc tagtgaccac tgctctcaca tttccccnaa
```

FIGURE 3 (continued)

```
65461 agaccagaac actgacttgg tggtatgggg gtgggaaaga tggtattctc atcactttca
65521 gttaaattat gaacattcag aaagctgcaa agtactcgca gcattaattt ctgtacagtt
65581 tgagatacag attgtactta gacttgattt ctttcaaatt gcaaatgtca gtgctctgtg
65641 aaggacaacc aagttccaaa tggggcagag gaggatttca ggggcctggt aaaagctagt
65701 gagtgacaca ggctatcagc taccaggatt gcagaaagga gttggggtga cagggactta
65761 cgttcatttt cctgcgacac cgagagaggt tgaggaggag agacaccttt agttcccgga
65821 aagttttcaa gtcctcacca aacccttctc tagggaactt cttcagggcg tactggtagc
65881 gctgggcagc ttcctttact ttacctttct aatgggagca caagatggcg aaacacacag
65941 ttggtctgct gaacagctga ggaagtgtgg tccccactcc cttcttcttc tcagcagctg
66001 ttactaaggg ctaaccctcc caaacgccag cctagagagg agttctggcc ctgatggtga
66061 gggcagaagg gtctagtctt ttctatagga aactcacctg gcaggaaagg caaggttgag
66121 gggataccca agatacaagc cagggagcaa agccttggca aatggaaggc cctggaccct
66181 gacttcccaa ctggattcca aagaggctga ccatgctcat ttcaaagttc ctttgctctc
66241 tataacacat atgacttcct aaccccaagc actccaaagt tattacttcc tccttaattt
66301 ctggcttgtc tggagaactg ttttttggt ctgacgctcc acctggccag ccctctagaa
66361 gtcactctct tcctcatgaa ggatgtggga gctaaagcag tctagcagtt gcttccctaa
66421 tctcatggca taaaataata ccaatcagtg ttccccattt ttggaaaggc atggagtagg
66481 gtggcagtag aaaagtgaag tttagagttt gtgattccat ctgatgttgc aggagggatt
66541 atccaagagg aagtaaatct atgaccaatt tagagatctg gccacagcga ccttgaactt
66601 ttcttccact ttgtatctga aaatctataa gctgagacca gaaagatgca ggcggcatct
66661 cccacccctt tagctgatct ttatccactc tattccttct ctatagttac aactccacag
66721 ggatcccagt cttttacttt tcccttagca ccattcacta aggagagatt tggagtcttt
66781 caaggtagaa ttctgagggg aacaataatg gcactgcata ttttcattaa gaggactcct
66841 cgcctcagaa taccagaagt tatgcnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
66901 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
66961 nnnnntttt ttttttttt ttttttttt tttttgtatt ttagtagaga cggggtttca
67021 ccgtgttacc caggctggtc gcgaactcct gagctcaggt aatccgcccg cctcggcctc
67081 ccaaagtcct gggattacag gcgtgacacc gcgcccggcc agttttactc ttcttaaatc
67141 tgatgttcat ggcataaagc aggagtagtg gctttaaata tcactagcca cagctttcag
67201 ttgctctgat cgctatcatt cacaactttg agtaaaagga agaaagaaaa ttcttccaac
67261 tgacttattc aactttcaaa aatgagtttg ttgcaatgcc agtttaaaag ctcccctact
67321 acaccgtgct tctataaagg attctcaaat tgatgtatgt ccagcaagac ttactactaa
67381 aggtcacaca actgtactag cacgaaacaa gccaagagat aaggccctgt catccttctg
67441 tgttcaggaa acaacggcgg gatcatgcca caggccaaga tttccaaaga gattgcttta
67501 gagagaatgg atatcaggaa gcaagagcaa cagtaattct ggcagaaggg ctaggatcca
67561 tgagtacccg ggggggtng ggagtagggg gcggggaagc ttgtacaata gatcgtctta
67621 atagaaatct agtgtggcca ggagtgcacg accaccaata gaagcgcctg cattttcctt
67681 cacagctatc actctctctt cccttccccg cccaccaacg aaagcaatgc ggcaggagtg
67741 gagagcacgg gatgaataaa tggagtatct ttggggcgta aacccggttt taagacctgt
67801 ttccaccccg aaatagctgc ctgacccagg caagtcacaa aaccgtgcct cagtttcccc
67861 actacactgc cgactacaca agaatgctgc ctggattaaa ggagacagac aacgagtacg
67921 aaaccttcag cacagggcca gacagaggaa ccgctccgtg gccctagggc gagtccgcaa
67981 gagggttctg aggcctgggc tccggcccgg aaangggcgc tcccggggcc cgctccccgc
68041 cagctgggtt ccgagcccct gcccggccca ccttgttgtt gtactcctcc acgaagctgc
68101 ccagcgccaa gcgaaagcgc ttnatcgggc cgcacactcc agttcatcgc gtagactgtc
68161 cagggcgctt catacttgta gatctccttc cgtttgccgt gcagggacat ggtggccacg
68221 gggncgcagt nacgggccgg gtcaacagtg ggctgcggcg ggcggggagn cgagcctgag
68281 atctatgggt ccgaagggag gagaggaggg ggcgggcagg aacggcctaa cccggaagcg
68341 gggatgcggc gacaaacaac gacgacggcc gagcccgacc cctagtttca aaccagcttg
68401 ggaacggacg acaaccacct tccgttttgg gacgccgccc cgccctctcg gaacggaagc
68461 cgccggaccc ccgcagcggc accggccgtt ggttgcctga cacgtccttt cgaaaggatt
68521 ctttcttgtc attggttatt gcggccgtag cgtctgaaat ctcctcgcta ttgggttaaa
```

FIGURE 3 (continued)

68581 tgtttgtcat tgcaactact cgccccgccc aactctcgga ggcaggcgct ttggcagccc
68641 caaaggtcat tggttggcca agatgtcagt caggcagata acggctcagt gcgggtggtg
68701 ggggcgtggg tggcctggag gcgagcgtgc tgtagcagcg ggcctccaag ttctaggcca
68761 agtctctgag agtgaaaccg tctgtgacct gcgctgactt cccctccgct gctgctgttc
68821 tgagcggcct ctccacgctg tcgagtaaaa gtacaattct gccttagtgg aagacctact
68881 gactttcgcg ggacagagcc ctgcggcctg gcagncctgg cctgcgccca acgggtacca
68941 ccttcccgcc ccatctcttc ccagggccct tttaacccga aagctgctct ctccgtgcct
69001 cagggacagg cttcgggcgc gcagatgcgc tcagggccag gccagcttga gtcaggccct
69061 cccgcgcttt cctgcaggat ctagaaatgt ccccgaattc tggggtaggc accgacccga
69121 ccagcttggc tctgtttttg tcagttccca tcctgtaccc ttcccgccgt ggatcccaac
69181 gcaaatgcta atccagcccc aaaggtatct gtgcctgtga gttggaaaag gggaagggcc
69241 agagaccaaa tgacaccttt atataaatta tttccaattc tcagctttca tcaaaattga
69301 aggaggtttg tattgaattg ttggatgata acatntaaat actttttttt ttttttttna
69361 aagacagggt cttgccctgt cacccaggct ggagtgcact ggtgtgataa gggctcactg
69421 tagcctggac tgcctaggtt caaacgattc ccccacttta gcctctggag tagctggaac
69481 cacagacatt tgccaccaca cctagctatt tttaaattat tatttgtaga gatgagatct
69541 ctccatgttg cccaggctga tctcgaactc ctgggctcaa gcgatcctcc cacatcagca
69601 tcctaaagta ctgggattac aggcttgagc caccattctg ggcactaaat ataatttta
69661 atgataaatg atttttccac ctcccaggtt caagtgattc tcctgcttca gcctcccgag
69721 tagctgagat tacaggggtg caccaccaca gccagctaat ttttgtattt ttagtagaga
69781 tggggtttca ccatgttggc cattctggtg tcgaactgct gacctcaaat gatcttccag
69841 cctcggcctc ccaaagtgtt gggattacag gcgtgagcca ccgcatctgg ccaagaaatc
69901 ctctttttt tttaatttta ttttattttt gagatggagt ctcactctgt cgcccgggct
69961 gtagtgcagt ggcactatct cggctcactg caacctctgc ctttagggtt caagtgattc
70021 tcttgcttca gcctcctgag tagctgggat tacaggcgcc cgccaccatg cccagctaaa
70081 ggaaatcccc tgtaaggagg aaagtttaaa aacaaattta aaaaggaaag aaaaagacgt
70141 acaaaattcg cggtacctgt cttattctaa taaaaaatat gatatttatc cacagatata
70201 ggaactaacg gggggagaaa attttctctt ttatgacgta tatttccaaa aaaaatttta
70261 gtttttatta gaaaattgaa tatatattta tgttgttttg aacaactgta ctactaatag
70321 atgtaaccct taaagcctta ttaaggttac aaaaattaat tttaaaatgt caattaattt
70381 atacacattc gttttgaggc agggtcttgc tctgttatcc aggtggtggt gcagtggggc
70441 aaacagtgct tattgcagcc ttgacctccc gggctcaagc aatcctcctg cctcctttat
70501 tttttgtaga gacaagatct tgctatgttg cccaggctgg gttcaactgg tcctcccacc
70561 tcagcctccc aaaatgctgg gattataggc atgagccacc ccatgaccga cctacttttt
70621 ttttgaaacg gagtcttgct ctgttgccca cgctggagtg cagtggcacg atcttggctc
70681 actacaagct ccacctcctg ggttcacgcc attctcctgc ctcagcctcc cgagtagctg
70741 ggactacagc acctgccacc acgcctggct atnnnnnnnn nnnnnnnnnn nnnnnnnnnn
70801 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
70861 nnnnnnnnnn nntgggttgt nttttcattt tctttctttt ttttttgag agggagtttt
70921 gctnttgttg cccaggcagg agtgaagtgg aacaatttgg gttnaaaaaa atttttgcct
70981 ccccggttca aggcaatccc ctcccccacc cccccgagta gctgggataa aaggaatgaa
71041 ccacaacgcc tggctaattt ttgtattttt agtagagatg gggtttctcc atattggcca
71101 gcctggtctc gaactcccga cctcaggcga tctacccacc tcggcttccc aaagtgctgg
71161 gattacaggt gtgagccact gtgcctggcc gtcttttcat tttcttgatg gtgtcattga
71221 agcacaaaag ttttaaattt tgatgaagac caatttatct gttttttctt tcatcactta
71281 tgcttttggt gtcatatcta agaaaccatt gactaatcca aggtcacaaa gatttattgc
71341 ctatgttttc ttctaaaagt tttatgattt tagttcttac atcaaggtct attttaagtc
71401 ttttgttttg ttttttgttt tttgttttga gacagggtct tactctgtca cccaggctgg
71461 agtgcagagg cacaatcatg gctcactgca gcctcaacct cttggcctca agcaatcctc
71521 ccacttcagc ctcccaagta gctggtatta tagacatgcg caaccatgcc cagctaattt
71581 ttttgtagag atagggtttc accatattgc ccagactggt ctcaaactcc taagctccag
71641 tgatccgccc acctcagctt cccaaagttc tgggattata ggcatgagcc actgcaccca

FIGURE 3 (continued)

71701 gccccaaatt ttgtatatgg tattagaaag gggtccaact tcattctttt acatgtggaa
71761 atccaattgc cccagcacca tttgttaaaa atattttctt tcccatttaa ttgtcctagt
71821 gttcttgtca aaaacaattg aacataattg tatgggttca tttctggact ctcaattcta
71881 ttccattgtt gagcatattt ttaagggctg tttttctctc ctgtggtaac tggtgacctg
71941 tacttcctgg aagagagatg aaaagattcc caagccaact gagttacctc acgtgggtca
72001 ggtctctgtg gctctctgca ctggcctatt cataatgata tcctcctgca ggatttgagc
72061 cttctccttt gttgtgacgg cagccgagga ggtggctgac tgcccagaca gccttatctc
72121 tttcctacct ttcaaggtta ttgtaaatac caaattagat tgtttataca caagaattta
72181 gcactaaagc actatacaaa tgtaagctat ttatttctat ttatccttct ccttcatgaa
72241 taagacccta aaaatagaag atatttttaa tttttactca ctgggctcaa ggttgcagtg
72301 tctgtattat tgcaaattcc aaattaatga agtctggctc ttcttatatt attcctgcaa
72361 aaggctgtgt gctcaccccc cggagtgtga atacgagtgt gggtcatttc ctctttcctc
72421 tgcacccttc cttcgatgag gttttgccct gtgctaggca ccatgctaaa ctctgagaaa
72481 accacagaga acaaggcaga cgccatccct gccctccagt aacatacaat ttagtgatga
72541 agctggggga tttaacaagc cattctaata aagcgttaca gtgagggagg tgcagggtga
72601 tgcctccagc agccctggtg tcagctgctc caggtccagg ggaagacttc cattatcttc
72661 caaccctgtc ggaagtggag gcggaggctg tttattgctg acacagtncc ctaacccagg
72721 gtctatagac attgtggaaa tgccttggag tcagacggga gaatgaacca gcagaagcaa
72781 tgcccgccct ccaccctcct gaagagggtt ctcaggaact ctttggaggc gaggccagcc
72841 tctggntgna gnggcctctg gatacaggtt aggcctcagg ctcttctcct ctctactcat
72901 ctctcctccc ttnggcccct ccttcagagg ctgacagagc cccactctca tctcttcccc
72961 acccaagcct ctttccacag aaagactgct tcctcccagg agacagcagc tcatttgcac
73021 acagacaccc acagccctca aagcctggaa ggccaagctg ttaggacccc tgagagcagg
73081 gtggctcctg ggaggagagc ccaggccacc accttgccct ccctggcccc tggccttncg
73141 atggggctgc tctgatcaca aacgtccacc aacgtagccg gcccagaagt gcacccatgt
73201 cctctggtat ccactggctc tccaagccaa actgggcagg gaggagttgt gagggaaaac
73261 tgcaggtcag gagggaggct ggcaaagcgg gccagggcca ggcctgaccc cagctctcct
73321 ctcccggccc cactgccggc cagtgtttaa caaggccctg ccttctccct ctagtgctag
73381 ggacagccac cttcttcctc tccccaccgc cccctctccc ctgcaacacg tcatctgaca
73441 agtcagtgcg atctcactgg aggtgcatct cacaggaacg cggggtcaca gcctcctgca
73501 cacactccat gctgcacagc aaggtgcacg tgtccctcag agccccagac accatccccc
73561 actcacccag aagcccaagt gattcccaac agcccccagc agcctaatgg gttggggtct
73621 tgggagcagc tgtccctggc tccttccctg atcccaccgc ccagcctcac cccacggttc
73681 ctccattgcc ccacctccca ctgcgccgcc gggcctctgc cagggtcaag gggcttcccc
73741 cctctggcag cagacgccat ggtgccgagg tggcctccac aaccgccctg tgcgccaata
73801 ggacaagacn tgtcctccct cccccacact tgtcactttg agggacacgt ggatgagaca
73861 ggaaaacaca ggggagtgtg gagacctgag gtgacttgga gcaagcctct caacctgagc
73921 ggcaatttct tcatctgtaa aatngagggg gttgttctca tctctgaggc tttgtgtcgc
73981 tctcaaagcc tgctagcctc gggttctagg actctgttgg gatcgtgtgt gatgttttct
74041 gctgagcgac tggcagcctg tgtcctcggg gggaaagagg gcaggcgctc caaagctcct
74101 gcgctctgtg gctcccccctc cctcgcagcc ccaagcccca ggtgtgccgg ccgccctgag
74161 cccctccagc acctcccgga ggcgcctgca agacacctaa ggtccccgcc tccctcctct
74221 ccccccccgcc acacccctac ccccggcagg cgacgtcccc gcccctcgac catggcctgg
74281 tgaagaagcc ggccaggccc gatcagcccc aaccccgccg cacgagcggc gnctgncgga
74341 cagctcctgg ggccccggcc ttgtcactcc ggaggcggga ggctccgggg ggtcgggctg
74401 ggaagatcga gccggaggcc gctaggctnc ccaggccccg gccganggct gcgcggccgc
74461 acggtgggca ggctcgggtg nttccggcaa actgccgggt ccccnatctt caaaagagag
74521 gaggcccttt ctccagcttc ctctgcggga gcccgaccca gccccatccc gccacccccg
74581 ggctgcacct cggccccntc ncnccggcnc ncgcgccccct gcccggggcg ggccagngaa
74641 ncctcggccc gcgccgctng gggactttgg agcggaggag gaagcgcggc ggggcggggg
74701 cgggggtgtg tcgggtttta ntaacccgca gggcggccgc ggcgcaggag aaggggcaga
74761 gccgagcacc gcgcaccgcg tcatgggggc cgcctcgggc cgccgggggc cggggctgct

FIGURE 3 (continued)

```
74821 gctgccgctg ccgctgctgt tgctgctngc cgccgcagcc cgccctggcg ttgngnaccc
74881 cgggctgcag cccngaaact tttctngctg acgaggccgg ggcgcagctc ttcgcgcaga
74941 gctacaactc cagcgccgaa caggtgctgt tccnagagcg tggccgccag ctgggcgcac
75001 gacaccaaca tcaccgcgga gaatgncnaa ggccaggtgg cgcccgggcc cgggcggggg
75061 cggggcgggc cgcggcggcc aatcacagca cgcggccggc ttgtggggnc gggcagnntg
75121 gcgcccccgn acccgaaccc caccccgacc ccggaccctc gccccgacag tcagccgcgg
75181 ggcccgagcg ccgggctgcg cgcacggcct gcgctcccag catgcacgag ttggatggat
75241 gagggtggct gctcccaggc cgcgcccgcc ttcgccgaag gtgctgggct tggctctggg
75301 gcccccgcgc tctcgggcag ctgccttctc acctccggac gctgtcgctg tcaccgtcac
75361 cgcactgcac tgtccatcca ccctccactc gcccggcctc ttttctggtc ccaatttctg
75421 ctccaccatt cccatgaggc agattccctn ccagaaggag gaagcgcggc ttccgcnaaa
75481 ctaaggtctc ccgcagggat nctccccagg gcccgggctc tggaagccct tggccttcct
75541 cccctccccc agcaccgtgg cttctccttt atggcctgca tctanagcag ggtcctacac
75601 cctccctgcc ctcctggtgc ccaataggag gaagcagccc tgctcagcca ggagttntgc
75661 nggaggcctn ggggcagnaa ggccaangga gctgtatgaa ccgatctggc agaacttcac
75721 ggnaccncgc angctgcgca ggatcatcgg agctgtgcgc accctgggct ctgccaacnc
75781 tgcccctggc taagcggcag canggntggg ctgagggctg aggcagagct cgggggcggc
75841 ctcctagtgc cccatcgtgg gggtcggggg agagcagccc atcangggag ggaggaaccc
75901 tgggatccac atgggcccnt gacagaaggg naaagcccag gtaagcacag aatggctttc
75961 tgagncattg atttttcttg gagatggggt ggggagttac tttctgttaa aggaagcatt
76021 cntggagtag gaagccaaat tcaaatacac ttctccctag gctggtttat gagcttcttt
76081 ggaagagttg agaagggctg ggcgtggtgg ctcaagcctg taatcccagc actttgggag
76141 gctgaggtgg gcgcatcgct tgagcccagg agttcaagac cagcctggcc aacatggcaa
76201 aacctcgtct ctacaaaaaa aaaatagctg ggcttggtgg tgcgtgcacc tacagtccca
76261 gctactcttg aaactgaggg ggaaggatca cctgagccca ggaggtcaag gctacagtga
76321 gctgtgattg cactactgca ccccagcctg cgtgacagag tgagacctcc ccccaaaaaa
76381 aagagagaga gaaaaggttg agaaagactg ggaagtcacc aaagccagag aatgggaggg
76441 atctgccctc actgcagggt ggtgccaagc tgggacttga ccctgaccct gactttcagg
76501 actcctgtcc cccactccac aggctgcctc cactggcagg ggactcagaa gtgatccggt
76561 cacactaagt gacacttagt gatcagaagt gccccggtgc cactgagtgg ccttgtccaa
76621 gctacatcca ctctgtgggc tcctccttgt agcagcgagg ggagggcaga tgtcccaggg
76681 gctggtcact ggagcattcc tcccctctga ctccccagta caacgccctg ctaagcaaca
76741 tgagcaggat ctactccacc gccaaggtct gcctccccaa caagactgcc acctgctggt
76801 ccctggaccc aggtacggcc cttgcagctc ccctctcggc ggtgccctag tgttcccaca
76861 ttgccctgct gcactccgga ccatgcagtt gtgtagggtc tgtggagaca gcaggtaaac
76921 ccaaaggtgt tgccctccaa ctggggctgg acggtgcaga tancccccac gccctgcttc
76981 tcttggcaag tggacttccg gaatctccag ctgcagcccc cacttctgtg tgtacctcgg
77041 cctctcccat cacccctagg ccttcctcct ggctgcctgg tttccccttt cgtgggtcct
77101 ctcatgttcc ccaagagccc tcaggccagg gacccctcgt ggcttccccc taaaccccgc
77161 tccagccccc tttatgagca gcttcgagga aggcactcca tccaataggc cgctaagtgt
77221 ctgtctgggt tttggccttt gggtgtcccc ttggtgtcag ccaccttagg tggtcatgtc
77281 tctggggcag gggccctgcc tggngtgttt ctgtagctcc cagccccctc ccaccaggcc
77341 tgtaggtggc ccctgtctct gggggcaccg tgatgttcag gaagctggtg ggagcagtaa
77401 ggactggtgc aggctctggt gaaggccgnt tgaagacttc aacgtggagg cctcctcacc
77461 gaccctgcct gcctgtgtct cagatctcac caacatcctg gcttcctcgc gaactacgnc
77521 catgctcctg tttgcctggg agggctggca caatgctgcg ggcatcccgc tgaaaccgct
77581 gtacgaggat ttcactgccc tcagcaatga agcctacaag caggacggtg agcaggcctc
77641 tccctgtcca ggaaccacgc caggtgtcct ctctcagctg tctccccaga gtcccagccc
77701 agagtcaggc agagcagctg gtatgacaat tccagcaggc cctgagtttc ccagaaagtg
77761 gaggtgggac cggcctgcac ccagtgtgcc tggactttgc tgctggcctg ccccacgtgg
77821 ccatcctgct gtcactcctg gccctgatgc tcctctttgc ctctgggaac ctccaggatc
77881 tgtttagctg gctgtagcta attagaaatt gtagagtggc aacccccaag ccaattttcc
```

FIGURE 3 (continued)

77941 agctagctgc agatccacgg gcctcgagcc agtggaagag ccgacttaca gctgagaggc
78001 tgaggtccga gcctttggcc tgagctacat acctcacccc cacgccccca ggcttcacag
78061 acacgggggc ctactggcgc tcctggtaca actcccccac cttcgaggac gatctggaac
78121 acctctacca acagctagag cccctctacc tgaacctcca tgccttcgtc cgccgcgcac
78181 tgcatcgccg atacggagac agatacatca acctcagggg acccatccct gctcatctgc
78241 tgggtaagga cctggcctcg cctccacatg agtcccacgg aagtgtgggt cccgaggtag
78301 gggtggggga tgtccagggt aagggaaggt gggttgtgac cctcacatct cacatgtgtg
78361 gggcatcata ctgtttgctt cacatgcagg agaccattcg tgttcccact ttacaggtgg
78421 ggaccctgag gcttagggtc gtgagggact tagtggtcag agagctaggg gccaaaccaa
78481 aggctctggc cctgggtcca gtgggggagc catcagccta gctcatgccc naaggaaaca
78541 agcactgtgg ccctgcctca ggattgagtg gctggggcct ggcacagcca gaaatgacag
78601 tggcagcatc ttgcagcccc aggacatgtg gccctcggag gagtgtgggt gggactgatg
78661 tgtgagattt ctggccctaa gccaggcctg ncagcccttg agggccccag ggtacaggtg
78721 ccggccccag ggtgccactc agcgatgcat gaagaagnca ggcacagcca ggcagggagc
78781 caagctgtcc ccttccttcc ttatctagga gacatgtggg cccagagctg ggaaaacatc
78841 tacgacatgg tggtgccttt cccagacaag cccaacctcg atgtcaccag tactatgctg
78901 cagcaggtaa gctctgggct caagcctngg ggtggtgggg gtcgggggtg gggcgcaaaa
78961 aaagggagtc acagatgggc acaggggcgg gaaggtttcg ggtactgagc agcagcctgg
79021 tgtgtctgta ggagcagtga gctggggtcg gccccctcag tgaggtgcca gctcctccct
79081 ccaggctcca cagtggcagg atgagagcaa caacgcactt tcactcatct gctgtgggag
79141 tgagggccct gcctctggga atggtggcca cagagcagag aagctttcat gcacagggag
79201 ttgacccgag atggggaccc cagccctgtc cccaggccag ccagagtggg ctccccctga
79261 cctggctcca cacccctcct ccagggctgg aacgccacgc acatgttccg ggtggcagag
79321 gagttcttca cctccctgga gctctccccc atgcctcccg agttctggga agggtcgatg
79381 ctggagaagc cggccgacgg gcgggaagtg gtgtgccacg cctcggcttg ggacttctac
79441 aacaggaaag acttcaggtt cagacatggg aagagcacgt tctggggttc cccggttctg
79501 gggcccgggg aaaggcaggc agcccaggcg cagggaagct ggttcccagg cctgcctcta
79561 ccctacccca gcactggttg gaggctgggt ctgttccagg gctagggggt ataggaggcc
79621 tattagtcca ccttctctgg cagctttgac aaatagtcac ttctatacct tggaatggag
79681 gaagaaggcc caagtggtgg tgagccaggg cagggtaaag aatttgcttg tttctgccag
79741 gcacggtggt cacacctgta atcccagcac tttgggaggt caaggcgggt ggatcacctg
79801 aggccaggag ttcgagacca gncctggcca acatggcgaa accccgtctc tactaaaaat
79861 acaaaaataa attagccagg ggtgatggcg ggcgcctgta atcccagcta ctcaggaggc
79921 tgaggcagga gaatctcttg aacccgggag gcggaggttg cagtgagctg agattgtgcc
79981 actgcaggcc agcctgtgca aaagagtgag acgctgtctc aaaaaaaaaa aaagaaaaaa
80041 agaagttact tgtttctact gcggcttcat gccccagggc agctccctcc tcattcctgt
80101 ctttcaggtg ccaatctgcc ctgtgccctg gccctgccct gttctgtcca tccgtcactc
80161 tcaccctcgc cctctctacg ccccaggatc aagcagtgca cacgggtcac gatggaccag
80221 ctctccacag tgcaccatga gatgggccat atacagtact acctgcagta caaggatctg
80281 cccgtctccc tgcgtcgggg ggccaacccc ggcttccatg aggccattgg ggacgtgctg
80341 gcgctctcgg tctccactcc tgaacatctg cacaaaatcg gcctgctgga ccgtgtcacc
80401 aatgacacgg gtatgggagg gctgagaggc ccccacccag cctcacctaa accccgctcc
80461 accccacagc aggacctcac ttgccccact cagctctgcc cttctttctg cctcccggcc
80521 ccaggtcagg cagggttcgg gatcctccta gagcctcacg gtgcacactg cgcccagctc
80581 agcacacctg ggggtcctct tccaagcagg gcccagggtc tcgagggcca gccatacctt
80641 ctctgcatct ccctggcctc actttctgct gccccgccag cccacactct taggggaccc
80701 tcttctccct ctgacctctt ccctctcctt tcatctcatc tcccaacaga aagtgacatc
80761 aattacttgc taaaaatggc actggaaaaa attgccttcc tgccctttgg ctacttggtg
80821 gaccagtggc gctgggggggt ctttagtggg cgtaccccca cttcccgcta caacttcgac
80881 tggtggtatc ttcggtgaga ggagggatag aaaagccttc gccccagcta gccctcccca
80941 gcctcctgga cagccaggcg cctcctgccc cagccagttc tagcctctcc tctctaatga
81001 tgtccccccgc tgtgacccac cgccttctcc tttcctgcct gaaactccct cttccaggaa

FIGURE 3 (continued)

81061 gtcttcccca gttcctcagg atggggaagg gttgccgggt ggaaatgcct tttctacaaa
81121 agctaaatcc atctgtttgc aacctctagg ccctaagaca atttaaccat ccttttccag
81181 aaccaagtat caggggatct gtcctcctgt tacccgaaac gaaacccact ttgatgctgg
81241 agctaagttt catgttccaa atgtgacacc atacatcagg tattagcgcc cccaccccac
81301 ccacccccag tactgtcaca ccctcaatcc acttctcctc ctgtgttcct agctgcctca
81361 tccccagggc ttgtcctcat gctcctccag acctcaaagg cctggagtta gagtggccca
81421 ctctcctgag cctgtcttgg gtctcccttc tcccccaaga tagcttctgg tccagcctct
81481 gccctgcagg aagctggatg gtgcctgggt aaggaacccc tgttcctggc cccccatgat
81541 cttccctgac tcccaccctg tgcctgcagg tactttgtga gttttgtcct gcagttccag
81601 ttccatgaag ccctgtgcaa ggaggcaggc tatgagggcc cactgcacca gtgtgacatc
81661 taccggtcca ccaaggcagg ggccaagctc cggtgtgtgg tgggaagccg ggggaagtgg
81721 gaggcagaga ggagcggctg gcaaagggtg tggcaggagg tgtctggctg ctctgatggg
81781 gtgggggca ccaaccacag agctggactg atgtggatgc ctgtctcctc gctatgtcat
81841 caaatattta ttgagtgggc cttctggctg gcantggggc gacacaaatg ccccctgcca
81901 ccatcagaga gatcccaggc cccagggtct tattgccaca gtttctgcag tccattgngg
81961 gggcggaagt ggccaggggc atgtgggccg gggtccagga gcagactcca gcctgagtcc
82021 cctgtgccca tggtacccac tctgcccacc aggaaggtgc tgcaggctgg ctcctccagg
82081 ccctggcagg aggtgctgaa ggacatggtc ggcttagatg ccctggatgc ccagccgctg
82141 ctcaagtact tccagccagt cacccagtgg ctgcaggagc agaaccagca gaacggcgag
82201 gtcctgggct ggcccgagta ccagtggcac ccgccgttgc ctgacaacta cccggagggc
82261 atnaggtaaa gccctgagtg aggatggtgt ggggctaagg tgggtcctca anctctgggc
82321 ttggcccagg nccccaggtt cctggtcagc tcctaccagc tgagccctgg taccctgtcc
82381 tggagggcca ggcagccccc caagctcatc agcagggcct gcgagtgggg acaggcatgt
82441 ctttccccca gcatcctaga gagggtgtgc tcagacctga gggcccctcc ccttccagag
82501 gaagccagac acaaggctct gtgaggtcac nactgcggcc tccgctcttn nnattggcca
82561 ggggacggta gctgcaggac tctgctctcc tgcggccatg ggccagggnt tgggctactg
82621 caggacttcc cagcctcctc ttcctgctgc tctgctacgg gcacccntct gctggtcccc
82681 agccaggagg catcccaaca ggtgacagtc acccatggga caagcagcca ggcaacaacc
82741 agcagccaga caaccaccca ccaggcgacg gcccaccaga catcagccca gagcccaagt
82801 gggaccatgc aggggagggg cagggtgcca ggggtgggag aggcggggcc gggntaggga
82861 cagggcaggg tacaagggag tgcgagaggg ataatggctt ctggtgagac cacaaacctg
82921 gagaggggag gcagaggttt gtctgtttcc ctgcactctg tcccacagac ctggtgactg
82981 atgaggctga ggccagcaag tttgtggagg aatatgaccg gacatcccag gtggtgtgga
83041 acgagtatgc cgaggccaac tggaactaca acaccaacat caccacagag accagcaaga
83101 ttctggtggg agccacctcc ccacccccaa acctgagcat gtgcatacac acagagatgc
83161 tgtcccgctc accacacagt ggggctgcca ccacatttta aattgaatat ttaaaacaat
83221 actcaatttc gggccgggcg cgggtgntca cgcctgtaat cccagcactt tgggaggggg
83281 aggcgggcgg atcacgaggt cagnnnatca agaccatcct ggctaacacg gtgaaacccc
83341 atctctagta aaaatacaaa aattagccgg gtgtggtggc gagcacctgt agtcccagta
83401 ctcaggaggc tgaggcagga gaatggcatg aacccgggag gcagagcttg cagtgagccg
83461 agatggcacc actgcactcc agcctgggcg acagagcgag actccatcaa aaaaaaaaaa
83521 aaaaaaaact caatttcaga ttttgatgaa catttactca atgcctgagc aattcttctt
83581 tccttaaaaa tcagtctctg ggaggcctag gtgggaggat cacttgaagc caggagttgg
83641 agactagcct gggcaacata gcaagatccc atctctattc aaacaaacaa ataaacaaaa
83701 atcaatctct agtaacagaa taatttgtac ataaataagt ggtgctcaag tcgtttttta
83761 aaagattgaa agcctctgtt tgtctcctct acaaaagggg ctacacttcc tctttaccct
83821 cattccctgc ctatttggct gagcacaaat tatgccactg agccacacac tgttactgtt
83881 ccttggcact ttgatctgtt gcctcatctt tttctcaaca gccttgcaaa attggtgagc
83941 ttattcccat tttacagatg ggatttgata ttaactctga ggttcagaaa ggccacagag
84001 ctaataccaa gctggctcct tcnctaaggg cctttaagac acttgggggt cttctcttct
84061 ctgcccctgc ctggatatgt gttgcttgac cgcaggcatc cagggagggt gagtactgca
84121 tccaggacgt tatcagcgtc cagcttgcag agagtcttat aggcaaaggt tgcaacttaa

FIGURE 3 (continued)

84181 ttccactgcc ccctcaccac cacctccagc cctcagctcc cacttggggc ctcccgctca
84241 gaggctgctc tggagctcct gggccctgtg acaccatccc cctgtgccct cagctgcaga
84301 agaacatgca aatagccaac cacaccctga agtacggcac ccaggccagg aagtttgatg
84361 tgaaccagtt gcagaacacc actatcaagc ggatcataaa gaaggttcag gacctagaac
84421 gggcagcact gcctgcccag gagctggagg aggtgtgtgg ctcgcaaggt acagggagag
84481 gggaatcctg gggcagtgag cccaacacag ggtctggcct ggccttcacg ctgcttcctc
84541 ttcctcgttg tatcaagtca tggcatctgc catgcgatng tgcacctcag aactgctgag
84601 agggcagcgc tccccagctc cctggctccc cacctgccag cccatggggc tngggggtag
84661 tgcaggcccc agagagacca agtgcaaagg agtacagctc attgcctctc cttcctcctg
84721 cagtacaaca agatcctgtt ggatatggaa accacctaca gcgtggccac tgtgtgccac
84781 ccgaatggca gctgcctgca gctcgagcca ggtgagagct catgtgcagg ctgagtgaga
84841 ggcgagggct gggactggca tggggcccgg gggtgctggg tgagagcaca gagttgggtt
84901 cccctcgctc ttggggtcag cgtgcccagg aaatgccctt tcttgttttc cacgaggggg
84961 gcttctctgc cccactgaga gccggcacct acttcatacc atgccccgat cagctgcccc
85021 tccctcagaa ccgccctctg cttaagggtg tccactctct cctgtcctct ctgcatgccg
85081 cccctcagag cagcgggatc tcaaagttat atttcatggg cttggactcc aaatgggggg
85141 aactcgggga cactagctcc ccccggcctc ctttcgtgac cctgcccttg acttcctcac
85201 cttctctgtc tttcctgagc ccctctccca gcatgtgact gataaggaaa ttgagtcaca
85261 cagcccctga aagcgccaga ctagaacctg agcctctgat tcctctcact tccctcacct
85321 accctgccac ttcctactgg atagaagtag acagctcttg actgtcctct tttctcccca
85381 ctggctggtc cttcttaccc cggcccgttt gaaagagctc acccccgaca caaggacccg
85441 cacacagata cctcccagct ccctctcaac ccacccttte cagggttgga gaacttgagg
85501 cataaacttg cttccatgag gaatctccac ccagaaatgg gtctttctgg cccccagccc
85561 agctcccaca ttagaacaat gacaaataga aggggaaatg gaaaatаaac aggagaaacg
85621 gttttcccag gacagggttt ggcctacaag ttgtggatgt gggtacccat gccaagtgtg
85681 aggggaggct ggccgggtgt ggtggctcat gcctctaatc ccagcacttt gggaggccaa
85741 ggtgagtaga tcacttgagg ccgggagttt gagaccagcc tggccaacat ggtgaaaccc
85801 catctgtact aaaaatacaa aagttagctg ggcgtggtgg tagatgcctg tagtcccagc
85861 tacttgggag gctgaggcat gagaatcgct tgagcccagc ctgggcaata cagcaagacc
85921 ccgtctctac aaataaaata caaaaaatta gttggatgtg gtggtgcatg cctgtagtcc
85981 tagctgctag ggaggctgag atggaaggat tgcttgagcc tgggaggtca aggctgcagt
86041 gagccgagat ggcgccactg cactccagcc tgggcaacag agtgagaccc tgtctcagaa
86101 aaaaaaaaaa aaaaaaaaag gagaggagag agactcaagc acgcccctca caggactgct
86161 gaggccctgc aggtgtctgc agcatgtggc cccaggccgg ggactctgta agccactgct
86221 ggagagccac tcccatcctt tctcccattt ctctagacct gctgcctata cagtcacttt
86281 tatgtggttt cgccaatttt attccagctc tgaaattctc tgagctcccc ttacaagcag
86341 aggtgagcta agggctggag ctcaaggcat tcaaacccct accagatctg acgaatgtga
86401 tggccacgtc ccggaaatat gaagacctgt tatgggcatg ggagggctgg cgagacaagg
86461 cggggagagc catcctccag ttttacccga aatacgtgga actcatcaac caggctgccc
86521 ggctcaatgg tgagtccctg ctgccaacat cactggcact tgggtccctt cattttcctc
86581 aaagaggtgc tgtgaaaccc caagcctagg aaaaggtaga tcccctggagg aggcaggtaa
86641 tgtggtgttg ggagagcctg gctgtgtccc ctctgtaggc tatgtagatg caggggactc
86701 gtggaggtct atgtacgaga caccatccct ggagcaagac ctggagcggc tcttccagga
86761 gctgcagcca ctctacctca acctgcatgc ctacgtgcgc cgggccctgc accgtcacta
86821 cggggcccag cacatcaacc tggaggggcc cattcctgct cacctgctgg gtaagggcac
86881 atgtcgggcc ttgaggaggg taaagacgga ccacagtgtg agtgagggtt gggacagggc
86941 tgactagagg gtagggagca ggctggggac tgagagactc cagccctgtg ggggatggtt
87001 gcccaggctg gaggggggtg ggcgctggga gtggggagcc ccccacttgc atctggtgcc
87061 acattcactg cagatctatg tcgggcaagt caccatggat gggggaagaa gttaataatc
87121 ttgtccagga gaccacggca cccatcacaa cattgtgtga tcttagaggg cgaggaagag
87181 gctgtgagtg ggagctgggg aggctttgcc aagaggtggc ctgtgagcag ggcctcggaa
87241 gatgacaggg tttgacagat gggaagtggg ggatgagagg acagacgcag tgttcaggcc

FIGURE 3 (continued)

```
87301 aagggaactg gaacaaagaa gaacctgaga atgtaaatct acttcaaccc tggaccctcc
87361 tttgccaagg gctgcaatct cagatgccct gaatgtgtga agtaggcggt gaggacagta
87421 agggatggta gggagtaagg caaagcagag gctactggtt ctctgtccct gatgggctgt
87481 taggaacact ttcctggagc agagagacca gacaggccct cagaccattt agaaactata
87541 agggaggccc cagaggacgg cctggctgtg ggtctagctc ccacacaggc tgggagtcca
87601 gccctcttca gcccctctct ggtgagacca aagaacatct ggtgatgtca cagtggacgt
87661 cagttcacca actgggagac acaggncccc gggaagaaaa gcaacatgcc cagcgtggcc
87721 tgggagctgg ggcagagctg gccttagaac tcagcccctg acaattggta aaaggggaaa
87781 ggggagcaac ctaacactga tgcgctctct gtctctctct ctggctctct ccctggctct
87841 ctccctcttc tctctcatgt tctctccatc actcatcgct ctaaccctct ctcactggtt
87901 tgcactttag acttcatctg atgtcagccg aagcttcacc tcacttggct aggaaagagc
87961 cttgagtcca aatctgtttc tgagccttcc attcatcctg agtttcttcc ttttcctctg
88021 tcgtggagaa ctaggctctt ttcttacact aaactcagag gcatcagcct ctnnctgaag
88081 gagacggctg gttcctgtca gagttgctga gctgcagaca ccgacctcag gtggtgcgga
88141 ggggacatgg cagagtggct ggtgaagaga gcagcctgcc agcctttcaa tcccagccct
88201 gccacttagg agccgtgggc ccccggcgag gggggcgnng tncacttaac tctccagcct
88261 gtttccttta ctagccaatg ggaatcgtga cagtacctgg gtgcagacag gattgaaagt
88321 gaattcacac aatgttcttg gtgcagagcc aataagaggt ggccaccggg gtgtaggtgt
88381 tctggggacc tgtaatgtcc tcacatgtca gcagttgcta gtcacattgg tctccactgc
88441 tcacggacag tgaagaccac ctggattccc tgataaccag caaggcnccc acctagagcc
88501 aggcagtaat gacctactgg gctggatatg cacaccaaag atgatgtgtg cctcaaagct
88561 tgcaaacagt aggtgctcaa gaaatgccac tatgattagc aggactggga tctggagcgc
88621 tcttcctgca ggagggcatt gagcctaagt aacatttgtc tttcctctct ctgccgtccc
88681 ccacactcgc ctccagggaa catgtgggcg cagacctggt ccaacatcta tgacttggtg
88741 gtgcccttcc cttcagcccc ctcgatggac accacagagg ctatgctaaa gcaggtccgc
88801 accagcccag gggcagggag gncccgccgg gantgggagg gaccctctga ttcaggagtt
88861 ccctccagtt tagccctccc ccgggatccc cacggcagca cgcagtctgn tccccggaac
88921 ccccagtttg ggcagaactc cctctngctt gcagggctgg acgcnccagg aggatgttta
88981 aggaggctga tgatttcttc acctnccctg gggctgctgc ccgtgcctcc tgagttctgg
89041 aacaagtcga tgctggagaa gccaaccgac gggcgggagg tggtctgcca cgcctcggcc
89101 tgggacttct acaacggcaa ggacttccgg tacatccagc tagggctcag gtctcgttcc
89161 tgagccccac gggcnaaggg aaatgaacca agcaaagggt ccactactgt cccccagctg
89221 gagccagcag ggcaggatgg ggacagggcc agagtttggg actgagtgtc tagagaggtg
89281 ttggcttctg gcaggaaaac cccatccgcc tgatggggac ttctgaagca cgcaacagct
89341 ctgtcagcct ggccgctggg aagtgctcaa ggtcccagtc ctgggtttga gcatggtagg
89401 ctgccccgcg tccctccttg ggagcagccc ctgcatggag ctggcctctc cctgggggca
89461 catgctgtga cacagggagg cacacgagga tgttgggtgc tctgtacaga tccactctca
89521 cccctgacag gctcagaagc tgccttcctt ggaggatggc gttttagtta cctattgctg
89581 tgcaaccaag caccccagag cttagcctta cgaaacaanc cagttgannt tttgcttatg
89641 attttatgtg ccaggaattc aggcagtaca cagtggaaat ggcctttctc tacagggccc
89701 cctctgttgg gggcagctct cacagccggg atggctcaat gggggccata cgtccagagc
89761 cccagttctg gctgtcggtt gaggtcctcg gtttttccca tgtggcagat gctggggcac
89821 atgttcccag tggcctcttg gctcacatgt tgggtgcttg gggtgagatg gctggaacgg
89881 ctggaggtgg gtcaggcatc tctccaggct agctcgggcg ccctcccagc gtggaatctg
89941 aggtgggcag atttacctgg cagccagcat cccccacagc aactactgac gcagccagtt
90001 ctcaaggcta ggtccaaaac tggcccagag tcacttctgc catgttttat tggctagaac
90061 aagtcacaag ttcacccaga ttcaagagaa gagaaaaaag tccctcccac ttgggagaag
90121 tggcaaagac catctgtcac agctgaagaa gtgtctctta caaggagaac agacacgggg
90181 agcctgaaac aaaacccgat gggattccct gggctgtgca ggcccttcca ggcatgagga
90241 ctcagccaca gggctngaga nggagacagg atctggggga tgagagccct tgtggggtct
90301 tccctttttat ggggagtcag aggagaagct ggatagatcc ccagccttgt ggccaggatg
90361 ctgggcagct cttccttccc cctccccgat gagaatgaca gaaaaacagg attcacctga
```

FIGURE 3 (continued)

```
90421 gccaaaaggc ttccagttag atccaagaga gaantttccc gcagtttgaa ttggtttgct
90481 aaacaacaag gaagggctgg gtgcggtggc tcacacctgt aatctcagca ctttgggaag
90541 ccgaggcagg aggtctgctt gagctcaggg gttcgagacc atcctgggca acatagcgag
90601 accccatttc ataaaaaata aataagtaaa tgagaacaag gaaggactga cgagagacgg
90661 tagaaccttc tggtttgggc nagctctgca gctgccattc atcctggcca taagaattct
90721 tggggtgaat aagtttgtcg ctgttgggcc gcatgagatg cagaatcgcc cactctcacc
90781 cctgacagaa acagttgttt ccttcaggga gcctccatct tgggagataa agcatgtgta
90841 catgggaacc cactggccac acattctcta gaaagtacac aatgtcccag tgcctctaga
90901 gcaagcactt tgtacagtca gaaagcaaca ggtggtgggg gctggagtca ttcaggaaaa
90961 tgggaggcag aggaatggcc tgaacggccc gatgctaggg gcttctgccc ccagattccc
91021 tcttacgcac actcagtggt tgcccttccc ctccctcccc acagtgctgt gtcccctgca
91081 tgctgcagtg ctggggtctg ccctgggtat agcaaggccc actgttccct tatgcccagg
91141 gcttctcact gtcctctccc aacaccctct cccccactcc actattccta ggatcaagca
91201 gtgcaccacc gtgaacttgg aggacctggt agtggcccac cacgaaatgg gccacatcca
91261 gtatttcatg cagtacaaag acttacctgt ggccttgagg gagggtgcca accccggctt
91321 ccatgaggcc attggggacg tgctagccct ctcagtgtct acgcccaagc acctgcacag
91381 tctcaacctg ctgagcagtg agggtggcag cgacggtgag agagaagcgg gaggccctgg
91441 tgggctgagg accaagaaag ggtggtgagc ttgggaggtg ggaaaggggc acttagtggc
91501 ccatgggcag aggtgtgggg cagagcaatc ggaaggaagg gagccaccca gaccatccca
91561 ggaggcaggt cacagggccc aaaaggtaca gcacccccac ccctccacca tcacaggcac
91621 accagggcca agccgctagg accctgggtc tgacagctgg gctcccttcc cttgcagagc
91681 atgacatcaa ctttctgatg aagatggccc ttgacaagat cgcctttatc cccttcagct
91741 acctcgtcga tcagtggcgc tggagggtat ttgatggaag catcaccaag gagaactata
91801 accaggagtg gtggagcctc aggttctgga acactcccac gggatgcggg ctgggggatc
91861 tctgcgagtg tctgcatgtg cctgggtgtc tggatgggcc agggtagggg agtgtgtgtg
91921 tgtgtgtaca ctattgtgtc tgtgcatatg atgtgtgtgg tgtaagtgat ggggaaaacg
91981 gggtgattgt gcacagaggc ccagcacgca ggagaatggg gtgcccagta tagccccaag
92041 tgcagggacc ctccctcaag tcaaaaatgc caccccagc ctggttctcc ccaaactcat
92101 cttccaacat atattcccac tcgacaggct gaagtaccag ggcctctgcc ccccagtgcc
92161 caggactcaa ggtgactttg acccaggggc caagttccac attccttcta gcgtgcctta
92221 catcaggtaa cgggaaaggc aggagggcac attgtgaggg gcagtaccca cagctttgtg
92281 tttcaactgc ggccactgcc cggtccacaa gctctgtcag tcagggcaga cccgggggag
92341 ccggccgcac ggtgcaggtg cctgggccca ctcacactgc caaggctgat gggtttttc
92401 ttgaacattc ttttgatgag agtctgtacc atccaaacag ttgaacacag aaactcaacc
92461 taataattgg ctaatggtta ccagaccttg gttaagtagt taacattaac cacgactcat
92521 ggctggatca tgagctctgc actgttttgt tttgctttta aaacaagact gtgattcttt
92581 tactattatt gaacattgtc tgcgatacaa tttgaattgt acctggaagc ccttctagac
92641 actaaaatgt aggattggag atcggttaag gtgggaggca gggttgctgg ggcaagttac
92701 agtcacaggc tggggtcaga cagaactggg ttcaaactct gtctccnatt actttgtttc
92761 cttgagaaaa ttcctcaatt tctgtgagct tccatttcct gacctgtgaa ccccatttca
92821 caggatgcac gatggctaac ttcttagcat tctgtctcat acacagcctc ctcagggagg
92881 ggtggccagg accccactat tcatcactct ctagtggaat gtagctgcac actaggtctg
92941 caggtcacat ggccacagat gagtgtgccc aatgcagccc ctctccttct gtgtgccccg
93001 ggagagcact tgctgagggc tagcaaggct gtttgtgatc cgggaggctc cctgggaggc
93061 tgggggctag agagacctca ggctggagtt ccaggtgccc ccgggctaca ggtagcccag
93121 gccaccccca gagggctgtg gctgcctctg gccctggcct cccgtggttc ctggaagccc
93181 agcaagggca gggcccatgc ccaccttgcc tcctggcacc tgggatgatg ccagcacatc
93241 atcaagtgct aataactgat tgtgggatgg atgaagtctg tccccagagt ccaggaagag
93301 ggcatccctg gagcacctca gataggncct gacctagacg gtgctccaga gatgacactt
93361 aggancaggg ctcccgcctg cctgctggag tggtccctgg ggttcccagc cggcgctggc
93421 tctcacccgg gagccagctg gtgtgatggc tagcttccca gcttaatgca gacaattctc
93481 caaacagggg tgggcaaagg agacttggct gctctagaaa aacattccgg attctggcca
```

FIGURE 3 (continued)

93541 gcagccttca caaagcactt ttaggaaaga ccagggaact aggtggtaca tgcttgcacc
93601 cagcattcaa agtgagaggc ctgtgccact ggctcaggac atttaaaacc tcttcagact
93661 ttaagctggg gagaatcctc cagccttgac tggcagattt ctaccaggga attcgtgatg
93721 ctttggataa gatcatgtag gactggcttc cctgcccaga ccaccctagt agatccacna
93781 cggcccattt ggccacacct tgcctctact tgcatatacc ccagggatgg agacctcact
93841 gcctccaaag ccacctgcca gatctcntga aggctctgcc ctgaccccat ggagcaggcc
93901 ctcctgagtt gtggaggcag ctctgtgggt gggaggcatc tacacaggca cggctaggaa
93961 gaggctcaga caatgctaag agctggggtg ggggagctca ccctgatagc tgtgggcaga
94021 gttggggggc cttggctctg ctgtgcgcat gtgacttagc acacatcaca tgtgatgtgc
94081 agaagggcct ggggcccagn tggcacaagg ccctcaacca actccgcccc gggccacggc
94141 ctcgctctgc tccaggtact ttgtcagctt catcatccag ttccagttcc acgaggcact
94201 gtgccaggca gctggccaca cgggcccccc gcacaagtgt gacatctacc agtccaagga
94261 ggccgggcag cgcctggcgn tgagtgtcct ccagccctcc tttgtttcca tngcntctgg
94321 cctgcgcccc tgggccttga ggggtctgtc cactggagct tttgtgggaa cacttgccat
94381 tttgagccgg gaactcccac ctgcagcgtg ggccaggcct gattgccatc tccttaggca
94441 cctggagccc tggggccctg ggacaagttt cagctgggag tgggtatgga gagtggatgt
94501 caggtggggg caagaggggc catgtccttc tgactctgcc tccctgtctc atgcctcccc
94561 aggaccgcca tgaagctggg cttcagntag gccgtggccg ganagccatg cagctgatca
94621 cgnnggccag cccnaacatg agcgcctcgg ccatgttgag ctacttcaag ccgctgctgg
94681 anctggctcc gcacggagaa cgagctgcat ggggagaagc tgggctggcc gcagtacaac
94741 tggacgnccg aactccggta ccgccaccca ccccacctcc agccttgggt cttaaccccc
94801 tccccaggct gggcagccat gcggctgacn ctncggagcc tggccctgcc ccgcaccctt
94861 gccctgccct gccctgccct gcccatgctg tctccttgct tcccgctcag ctcgctcaga
94921 agggcccctc ccagacagcg gccgcgtcag cttnnctggg cctggacctg gatgcgcagg
94981 aggcccgcgt gggcncagnt ggctgctgct cttcctgggc atcgccctgn ctggtagcca
95041 ccctgggcct cagccagcgg ctncttncag catccgccac cgcagcctcc accggcactc
95101 ccacgggccc cagttcggct ccgaggtgga gctgagacac tcctgaggtg acccggctgg
95161 gtcggccctg cccaagggcc tcccaccaga gactgggatg ggaacactgg tgggcagctg
95221 aggacacacc ccacacccca gcccaccctg ctcctcctgc cctgtccctg tccccctccc
95281 ctcccagtcc tccagaccac cagccgcccc agccccttct cccagcacac ggctgcctga
95341 cactgagccc cacctctcca agtctctctg tgaatacaat taaaggtcct gccctcccca
95401 tctgagtctg tgtccctcac agggaagcca gggacaggga caggctgctt tcctgcctcc
95461 tggcagtcaa gtgggtcccg ttactaggtt tgttcctcca tcctccttca ggagccgggg
95521 aggatcccca gagctctgcc ccagcacctn cctggcntgg cgcctgnntc ttccctccag
95581 cccaggcagc ccgccactgt cctgccaccg caggcagccc ctgtctggcc caagcactga
95641 cccacgcgga ctctgggaag cagacatcct gggctgctgg cctcacattt ccactggcag
95701 tggagccttt ccctgctcca caaatggcca ggtcccccca ggggaaggct tccggctgtt
95761 atcggctgcc tcaggggcg agtaccttgg agggcctgct tcaanggagg gtgccccctg
95821 gagggcacac accagcctag tgcttacctt ggctcctgcc tgtaccagct ccatgactct
95881 gctcgggtga acagccttgg ctctcagaca gccattctaa cactgccagt gcagaggggc
95941 ctcagacgct ggagtgtagc agtggctgca cctgcacagg gattagctgc cagcagccac
96001 cctgctggcg tcccagcaca cacctcctca ctccctgcat tggagggagt gtcatttaa
96061 gggacatttt tatgactttt atgtgtatgt ttatgtagaa atttggaaaa tacagaaaac
96121 tgtaaagaaa ataaaagccc tttatatcaa cgtcaagaga taagccctgt tgacgttttg
96181 gtgtacaact tgccggactt cttctcagca catgtgtatt ttaaatggga tcacaccata
96241 tttacagtca tacatccttt tatcacttca tacaactaga tctgtttttc tgatatttaa
96301 atgccagact tcgaacttgg ccaatagaag ttggtccatt tgctggggcc aggggctccc
96361 cagccacccg ggcccctctg tcaaaccctc agccctgagt ctcttctggg ctttgctgat
96421 gtcttcaccc tagctaggtc catctcccaa tatctgtccc ccttagtcca cagcttttgc
96481 cccccaatcc aggtgccgtg cgtgtctctg tgtgtccgtg tctgtgtgtg cgttgtacac
96541 aggcttggct gttacaggcc cattctgtaa ggcaggatgt ggggctgagg tattttagga
96601 ttgaaagagg gtggaagttt atgattacat aggacaatgg aatttataaa catgtcctct

FIGURE 3 (continued)

```
96661 aaatggttcc gagtcatcta ccaatgaaga cttcttgaat tatccccccc tttcccagcc
96721 tgttttgaaa gctctttgtt taccagacaa aggtcatcaa tcatatgacc ccctttgcct
96781 ttttttttt tttaagatgg agtctcgctc ttgttgccca ggctagagtg cagtggtgtg
96841 atctcggctc actgtaacct ccacctcctg gctttcaagc gattctcctg cctcagcctt
96901 ccgagtagct gggattacag gcgcccacca ccatgcctgg ctaaattttt gtatttttag
96961 tagagatggg gtttaatcat gttggccagg ctggtctcga attcctgacc tcaggtgatc
97021 cacccacctc ggcctcccaa agtgctggaa ttacaggcat gagccaccat gcctggcccc
97081 cgtttcctat ttttatgaac cacagcggtt catgctgcct gtcagagctt ctgggccgcg
97141 tgaggtcacc agctttcaac acgcaaagga ctgcactgca gctgggggaa gagaaactcc
97201 acactgcatt ggcctggcca gccttaccct ctgggctttt gaaatagtat ctttttctg
97261 tttgttttca aacagagtct cgctctgtcg cccaggctgg agtgctggag tgcagtggcc
97321 tgatctcggc tcactgcaac ctccacctcc caagttcaag cgattctcct gcctcagcct
97381 cccgagtagc tgggctacta cttncaggcg cacgccgccn atgcccagct aatttctttt
97441 gtattttagt agagatggag tttcaccatg ttgcccaggc tggtcttgaa ctcctgagct
97501 caggcaatcc gcccgcctca gcctctcaaa gtgctaggat tancaggcgt gagccaccgc
97561 gcccggccca atagtatcat tctttagatg cctgcctctg cctccttggg tgagtgggga
97621 gaggcagggg atacctggaa agtagcagag gaagaggagg cggtaacagc aggaagaggg
97681 ccagcccagt gttttctact ggtggccctg aaggctgagc ccatccccgt gccgtgcctg
97741 ccaatgccgc tcttgggaga ccagctctca cctacgctag ccacaggtgg tggctgccag
97801 acagtttctc tgatccccac agccctcccc accctctacc ttcctctgtc tgcctaaccc
97861 ccttcccacc caccctggct tttaacataa gtgaaaaagt ggctaacccc acctctgcac
97921 ttatcacctg tgtgaccttg ggcagtttgt ttttgcagtc tgcattttct tttcttttct
97981 ttttttttt ttttttttt tgagatggag tctcgctctg tcacccaggc tggagtgcag
98041 tggcgtgatc tcggctcacc gcaagcttgg cctcctgggt tcacgccatt ctcctgcctc
98101 agcctcccga gtagctggga ctacaggcgc cagccaccac gcccggttaa ttttttgtat
98161 ttttagtaga gacagggttt caccgtgtta gccaggatgg tctcaatctc ctgacctcgt
98221 gatttgcctg cctcggcctc ccagagtgct gggattacag gcgtgagcca ccgcgcccag
98281 cctgcgtttt ctttcttacg gttcttatca accctctcag cgttgccatg aagatgaaat
98341 gagatgatgt acaaagtcct agtagagtgt cttctcttta taatgaatgc atcgtctcct
98401 gagaaagcta gtttcataac aaccccagat cagccaagtc cagatcagcc ctctcacttg
98461 agacaggaag aggacccggg gcaactgggt gccggagctg gactgaaaac tcccatctcc
98521 cagctgcctt ccaggaactt ccccaccaca cgtccttgca caaccagtca actgtcttct
98581 tactgggagc acacagagct cgtccactgg gggcccacag cttgcctcag ttccaggagt
98641 actcagccat ctccctgtgt cgcctctccc tcatcatccc tccccatgtc acatctccct
98701 cagcctctcc gtgttacctc tccctcatcc tccctcccca tgttgcctct ccctcatcct
98761 ctctctcatc ctccctgccc gtgtcgcctc ttcctcatcc tctccctcat cctctccctc
98821 ttcctctct
```

FIGURE 4

```
1    mgaasgrrgp glllplplll llppqpalal dpglqpgnfs adeagaqlfa qsynssaeqv
61   lfqsvaaswa hdtnitaena rrqeeaalls qefaeawgqk akelyepiwq nftdpqlrri
121  igavrtlgsa nlplakrqqy nallsnmsri ystakvclpn ktatcwsldp dltnilassr
181  syamllfawe gwhnaagipl kplyedftal sneaykqdgf tdtgaywrsw ynsptfeddl
241  ehlyqqlepl ylnlhafvrr alhrrygdry inlrgpipah llgdmwaqsw eniydmvvpf
301  pdkpnldvts tmlqqgwnat hmfrvaeeff tslelspmpp efwegsmlek padgrevvch
361  asawdfynrk dfrikqctrv tmdqlstvhh emghiqyylq ykdlpvslrr ganpgfheai
421  gdvlalsvst pehlhkigll drvtndtesd inyllkmale kiaflpfgyl vdqwrwgvfs
481  grtppsrynf dwwylrtkyq gicppvtrne thfdagakfh vpnvtpyiry fvsfvlqfqf
541  healckeagy egplhqcdiy rstkagaklr kvlqagssrp wqevlkdmvg ldaldaqpll
601  kyfqpvtqwl qeqnqqngev lgwpeyqwhp plpdnypegi dlvtdeaeas kfveeydrts
661  qvvwneyaea nwnyntnitt etskillqkn mqianhtlky gtqarkfdvn qlqnttikri
721  ikkvqdlera alpaqeleey nkilldmett ysvatvchpn gsclqlepdl tnvmatsrky
781  edllwawegw rdkagrailq fypkyvelin qaarlngyvd agdswrsmye tpsleqdler
841  lfqelqplyl nlhayvrral hrhygaqhin legpipahll gnmwaqtwsn iydlvvpfps
901  apsmdtteam lkqgwtprrm fkeaddffts lgllpvppef wnksmlekpt dgrevvchas
961  awdfyngkdf rikqcttvnl edlvvahhem ghiqyfmqyk dlpvalrega npgfheaigd
1021 vlalsvstpk hlhslnllss eggsdehdin flmkmaldki afipfsylvd qwrwrvfdgs
1081 itkenynqew wslrlkyqgl cppvprtqgd fdpgakfhip ssvpyiryfv sfiiqfqfhe
1141 alcqaaghtg plhkcdiyqs keagqrlata mklgfsrpwp eamqlitgqp nmsasamlsy
1201 fkplldwlrt enelhgeklg wpqynwtpns arsegplpds grvsflgldl daqqarvgqw
1261 lllflgiall vatlglsqrl fsirhrslhr hshgpqfgse velrhs
```

FIGURE 5

```
1    gaattcatgc ccctttgaa atagacttat gtcattgtca gaaaacataa gcatttatgg
61   tatatcatta atgagtcacg attttagtgg ttgccttgtg agtaggtcaa atttactaag
121  cttagatttg ttttctcaca tattctttcg gagcttgtgt agtttccaca ttaatttacc
181  agaaacaaga tacacactct ctttgaggag tgccctaact tcccatcatt ttgtccaatt
241  aaatgaattg aagaaattta atgtttctaa actagaccaa caaagaataa tagttgtatg
301  acaagtaaat aagctttgct gggaagatgt tgcttaaatg ataaaatggt tcagccaaca
361  agtgaaccaa aaattaaata ttaactaagg aaaggtaacc atttctgaag tcattcctag
421  cagaggactc agatatatat aggattgaag atctctcagt taagtctaca tgaaaaggat
481  ggtttcttgg agcttccaca aacttaaaac catgaaacat ctattattgc tactattgtg
541  tgtttttcta gttaagtccc aaggtgtcaa cgacaatgag gaggtgaatt ttttaaagca
601  ttattatatt attagtagta ttattaatat aagatgtaac ataatcatat tatgtgctta
661  ttttaatgaa attagcattg cttatagtta tgaaatggaa ttgttaacct ctgacttatt
721  gtatttaaag aatgtttcat agtatttctt atataaaaac aaagtaattt cttgttttct
781  agtttatcac ctttgttttc ttaagatgag gatggcttag ctaatgtaag atgtgttttt
841  ctcacttgct attctgagta ctgtgatttt catttacttc tagcaataca ggattacaat
901  taagaggaca agatctgaaa atctcacaaa ctataaaata ataaaagagc agaatttaa
961  gataaaagaa actggtggta ggtagattgt tctttggtga aggaaggtaa tatatattgt
1021 tactgagatt actatttata aaaattataa ctaagcctaa aagcaaaata catcaagtgt
1081 aatgatagaa aatgaaatat tgctttttc agatgaaaag ttcaaattag agttagtgtg
1141 tattgttatt attaatagtt atgaaacacg gttcagtcta atttatttat ttgtagaaca
1201 gtttgtcctc aactattatt tttgctgact tattgctgtt aatttgcagt tactaaaaat
1261 acagaaatgc atttaggaca atggatattt aagaaattta aattttatca tcaaacgtat
1321 catggccaaa tttcttacat atagcatagt atcattaaac tagaaataag aatacacaat
1381 aatatttaaa tgaagtgatt catttcggat cattattgag tttcaaggga acttgagtgt
1441 tgtacttatc agactctaca tgtaagaaca tatagttaat ctggttgtgt gtgtaaaaac
1501 atatggttaa tctggttaag tctggttaat catattaggt aagaaaaatg taaagaatgt
1561 gtaagacgaa atttttgtaa agtactctgc aaagcacttt cacatttctg cttatcaact
1621 aaacctcaca gagatagttt aatagtttag gctttaaaat ggattttgat tattcaacaa
1681 gtggccttca taatttcttt aagtgttttt ctttaagtat atactttctt taaatatttt
1741 ttaaaatttc cttttctcta gtaaagccag accatccatg ctacctctct agtggcactc
1801 tgaaataaaa agaaaatagt tttctctgtt ataattgtat ttgtaataag cagatgaatc
1861 acatttctta aaatttgttt tagagagggt aagctctgac taggaccatg acttcaatgt
1921 gaaatatgta tatatcctcc gaatctttac atattaagaa tgtatatagt caactggtta
1981 aacaggaaaa tctggaacag cctggctggg ttttaatctt agcaccatcc tactaaatgt
2041 taaataatat tataatctaa tgaataaatg acaatgcaat tccaaataga gttcatctga
2101 tgacttctag actcacaaaa ttgcaagaga gctcagttgt tgctcagttg ttccaaatca
2161 tgtcgtttgt taatttgtaa ttaagctcca aaggatgtat agctactgac aaaaaaaaaa
2221 atgagaatgt agttaatcca aatcaaaact ttcctattgc aatgcgtatt ttctgcttca
2281 ttatccttta atataatatt ttaagttagc aagtaatttt aattacaatg cacaagcctt
2341 gagaattatt ttaaatataa gaaaatcata atgtttgata aagaaatcat gtaagaaatt
2401 tcaagataat ggtttaacaa ataattttgt tgatagaaga taagactaaa agtgaaattc
2461 gaagtggaga ggacacttaa actgtagtac ttgttatgtg tgattccagt aaaaatagta
2521 atgagcactt attattgcca agtactgttc tgagggtacc atatgcaata agttatttaa
2581 tccttacaat aatcttgtaa ggcagattca aactatcatt acacttattt tacagatgag
2641 aaaactgggg cacagataaa gcaacttgcc caaggtctca tagctgtaag tcaaccctac
2701 ggtcaagacc tacaagtagc cgagctccag agtacattat gagggtcaaa gattgtctta
2761 ttacaaataa attccaagta gaatcaacct ttaataagtc tttaatgtct cttaaatatg
2821 tttatatagg agtctaatca ccaattcaca aaaatgaaag tagggaaatg attaacaata
2881 atcataggaa tctaacaatc caagtggctt gagaatattc attcttcttg acagtataga
2941 ttctttacaa tttcgtaagt tccaatgtat gttttaggaa tatgaggtca ttactattca
3001 taatctgata cagctttatc ctaaggcctc tctttaaaaa ctacactgca tcatagcttt
```

FIGURE 5 (continued)

3061 tttgtgcagt tggtctttct actgttactg aacagtaagc aacctacaga ttcactatca
3121 ccaaccagcc agttgatgga tcttaagcaa attatcaagc ttgtgataac ctaaattata
3181 aaatgagggt gttggaatag ttacattcca aatcttctat aacactctgt attatatttc
3241 tgcctcattc cttgtagggt ttcttcagtg cccgtggtca tcgacccctt gacaagaaga
3301 gagaagaggc tcccagcctg aggcctgccc caccgcccat cagtggaggt ggctatcggg
3361 ctcgtccagc caaagcagct gccactcaaa agaaagtaga aagaaaagcc cctgatgctg
3421 gaggctgtct tcacgctgac ccagacctgg tgggtgcact gatgtttctt gcagtggtgg
3481 ctctctcatg cagagaaagc ctgtagtcat ggcagtctgc taatgtttca ctgacccaca
3541 ttaccatcac tgttattttg tttgtttatt ttggaaataa aattcaaaac ataaacatat
3601 tgggcctttg gtttaggctt tctttcttgt tttctttggt ctgggcccaa aatttcaaat
3661 taggatatgt gggtgccacc tttccatttg tattttgcca ctgcctttgt ttagttggta
3721 aaattttcat agcccaatta tatttttttct ggggtaagta atattttaaa tctctatgag
3781 agtatgatga tgactttcga atttctggtc ttacagaaaa ccaaataata aatttttatg
3841 ttggctaatc gtatcgctga attttcctat gtgctatttt aacaaatgtc catgacccaa
3901 atccttcatc taatgcctgc tattttcttt gtttttaggg ggtgttgtgt cctacaggat
3961 gtcagttgca agaggctttg ctacaacagg aaaggccaat cagaaatagt gttgatgagt
4021 taaataacaa tgtggaagct gtttcccaga cctcctcttc ttcctttcag tacatgtatt
4081 tgctgaaaga cctgtggcaa aagaggcaga agcaagtaaa aggtagatat ccttgtgctt
4141 tccattcgat tttcagctat aaaattggaa ccgttagact gccacgagaa tgcatggttg
4201 tgagaagatt aacatttctg ggttagtgaa tagcattcat acgcttttgg gcaccttccc
4261 ctgcaacttg ccagataagc actattcagc tcttattccc agtctgacat cagcaagtgt
4321 gattttctat gaaaaattct actatgactc cttattttaa gtatacaaga aacttgtgac
4381 tcagaagata atatttacag agtggaaaaa aacccctagc atttatagtt ttaacatttg
4441 aggttttgaa tgagagagtt atccataata tattcaattg tgttgtggat aatgacacct
4501 aacctgtgaa tcttgaggtc agaatgttga gtgctgttga cttggtggtc aggaaacagc
4561 tagtgcgtga gcctggcaca ggcatctcag tgagtagcat acccacagtt ggaaattttt
4621 caaagaaatc aaaggaatca tgacatctta taaatttcaa ggttctgcta tacttatgtg
4681 aaatggataa ataaatcaag catatccact ctgtaagatt gaacttctca gatggaagac
4741 cccaatactg ctttctcctc ttttccctca ccaaagaaat aaacaaccta tttcatttat
4801 tactggacac aatctttagc gtatacctat ggtaaattac tagtatggtg gttaggattt
4861 atgttaattt gtatatgtca tgcgccaaat catttccact aaatatgact atatatcata
4921 actgcttggt gatagctcag tgtttaatag tttattctca gaaaatcaaa attgtatagt
4981 taaatacatt agttttatga ggcaaaaatg ctaactattt ctacataatt tcatttttcc
5041 agataatgaa aatgtagtca atgagtactc ctcagaactg gaaaagcacc aattatatat
5101 agatgagact gtgaatagca atatcccaac taaccttcgt gtgcttcgtt caatcctgga
5161 aaacctgaga agcaaaatac aaaagttaga atctgatgtc tcagctcaaa tggaatattg
5221 tcgcacccca tgcactgtca gttgcaatat tcctgtggtg tctggcaaag gtaactgatt
5281 cataaacata ttttagaga gttccagaag aactcacaca ccaaaaataa gagaacaaca
5341 acaacaacaa aaatgctaag tggattttcc caacagatca taatgacatt acagtacatc
5401 ataaaaatat ccttagccag ttgtgttttg gactggcctg gtgcatttgc tggttttgat
5461 gagcaggatg gggcacaggt agtcccaggg gtggctgatg tgtgcatctg cgtactggct
5521 tgaacagatg gcagaaccac agatagatgt agaagtttct ccattttgtg tgttctggga
5581 gctcatggat attccaggac acaaaaggtg gagaagagct ttgttcatcc tcttagcaga
5641 taaacgtcct caaaactggg ttggacttac taaagtaaaa tgaaaatcta atatttgtta
5701 tattattttc aaaggtctat aataacacac tccttagtaa cttatgtaat gttatttaa
5761 agaattggtg actaaataca aagtaattat gtcataaacc cctgaacata atgttgtctt
5821 acatttgcag aatgtgagga aattatcagg aaaggaggtg aaacatctga aatgtatctc
5881 attcaacctg acagttctgt caaaccgtat agagtatact gtgacatgaa tacagaaaat
5941 ggaggtaagc tttcgacagt tgttgacctg ttgatctgta attatttgga taccgtaaaa
6001 tgccaggaaa caaggccagg tgtggtggct catacctgta attccagcac cttgggaggc
6061 caaagtgggc tgatagcttg agcctaggag tttgaaacta gcctgggcaa cataatgaga
6121 ccctaactct acaaaaaaaa aaaaaatacc aaaaaaaaaa aaaaaatcag ctgtgttggt

FIGURE 5 (continued)

6181 agtatgtgcc tgtagtccca gctatccagg aggctgagat gggagatcac ctgagcccac
6241 aacctggagt cttgatcatg ctactgaact gtagcctggg caacagagga tagtgagatc
6301 ctgtctcaaa aaaaaaaatt aattaaaaag ccaggaaaca agacttagct ctaacatcta
6361 acatagctga caaaggagta atttgatgtg gaattcaacc tgatatttaa aagttataaa
6421 atatctataa ttcacaattt ggggtaagat aaagcacttg cagtttccaa agattttaca
6481 agtttacctc tcatatttat ttccttattg tgtctatttt agagcaccaa atatatacta
6541 aatggaatgg acaggggatt cagatattat tttcaaagtg acattatttg ctgttggtta
6601 atatatgctc tttttgtttc tgtcaaccaa aggatggaca gtgattcaga accgtcaaga
6661 cggtagtgtt gactttggca ggaaatggga tccatataaa cagggatttg gaaatgttgc
6721 aaccaacaca gatgggaaga attactgtgg cctaccaggt aacgaacagg catgcaaaat
6781 aaaatcattc tatttgaaat gggatttttt ttaattaaaa aacattcatt gttggaagcc
6841 tgttttaggc agttaagagg agtttcctga caaaaatgtg gaagctaaag ataagggaag
6901 aaaggcagtt tttagtttcc caaaatttta tttttggtga gagattttat tttgtttttc
6961 ttttaggtga atattggctt ggaaatgata aaattagcca gcttaccagg atgggaccca
7021 cagaactttt gatagaaatg gaggactgga aaggagacaa agtaaaggct cactatggag
7081 gattcactgt acagaatgaa gccaacaaat accagatctc agtgaacaaa tacagaggaa
7141 cagccggtaa tgccctcatg gatggagcat ctcagctgat gggagaaaac aggaccatga
7201 ccattcacaa cggcatgttc ttcagcacgt atgacagaga caatgacggc tggtatgtgt
7261 ggcactcttt gctcctgctt taaaaatcac actaatatca ttactcagaa tcattaacaa
7321 tatttttaat agctaccact tcctgggcac ttactgtcag ccactgtcct aagctcttta
7381 tgcatcactc gaaagcattt caactataag gtagacattc ttattctcat tttacagatg
7441 agatttagag agattacgtg atttgtccaa tgtcacacaa ctacccagag ataaaactag
7501 aatttgagca cagttacttt ctgaataatg agcatttaga taaataccta tatctctata
7561 ttctaaagtg tgtgtgaaaa ctttcatttt catttccagg gttctctgat actaagggtt
7621 gtaaaagcta ttattccagt ataaagtaac aaacacagtc cctagatgga ttgccacaaa
7681 ggcccagtta tctctctttc ttgctatagg gcacaggagg tctttggtgt attagtgtga
7741 ctctatgtat agcacccaaa ggaaagacta ctgtgcacac gagtgtagca gtcttttatg
7801 ggtaatctgc aaaacgtaac ttgaccaccg tagttctgtt tctaataacg ccaaacacat
7861 tttctttcag gttaacatca gatcccagaa aacagtgttc taaagaagac ggtggtggat
7921 ggtggtataa tagatgtcat gcagccaatc caaacggcag atactactgg ggtggacagt
7981 acacctggga catggcaaag catggcacag atgatggtgt agtatggatg aattggaagg
8041 ggtcatggta ctcaatgagg aagatgagta tgaagatcag gcccttcttc ccacagcaat
8101 agtccccaat acgtagattt ttgctcttct gtatgtgaca acatttttgt acattatgtt
8161 attggaattt tctttcatac attatattcc tctaaaactc tcaagcagac gtgagtgtga
8221 ctttttgaaa aaagtatagg ataaattaca ttaaaatagc acatgatttt ctttttgttt
8281 cttcatttct cttgctcacc caagaagtaa caaaagtata gttttgacag agttggtgtt
8341 cataatttca gttctagttg attgcgagaa ttttcaaata aggaagaggg gtcttttatc
8401 cttgtcgtag gaaaaccatg acggaaagga aaaactgatg tttaaaagtc cacttttaaa
8461 actatattta tttatgtagg atctgtcaaa gaaaacttcc aaaaagattt attaattaaa
8521 ccagactctg ttgcaataag ttaatgtttt cttgttttgt aatccacaca ttcaatgagt
8581 taggctttgc acttgtaagg aaggagaagc gttcacaacc tcaaatagct aataaaccgg
8641 tcttgaatat ttgaagattt aaaatctgac tctaggacgg gcacggtggc tcacgactat
8701 aatcccaaca ctttgggagg ctgaggcggg cggtcacaag gtcaggagtt caagaccagc
8761 ctgaccaata tggtgaaacc ccatctctac taaaaataca aaaattagcc aggcgtggtg
8821 gcaggtgcct gtaggtccca gctagcctgt gaggtggaga ttgcattgag ccaagatc

FIGURE 6

```
1   mkrmvswsfh klktmkhlll lllcvflvks qgvndneegf fsarghrpld kkreeapslr
61  papppisggg yrarpakaaa tqkkverkap daggclhadp dlgvlcptgc qlqeallqqe
121 rpirnsvdel nnnveavsqt ssssfqymyl lkdlwqkrqk qvkdnenvvn eysselekhq
181 lyidetvnsn iatnlrvlrs ilenlrskiq klesdvsaqm eycrtpctvs cnipvvsgke
241 ceeiirkgge tsemyliqpd ssvkpyrvyc dmntenggwt viqnrqdgsv dfgrkwdpyk
301 qgfgnvatnt dgknycglpg eywlgndkis qltrmgptel liemedwkgd kvkahyggft
361 vqneankyqi svnkyrgtag nalmdgasql mgenrtmtih ngmffstydr dndgwltsdp
421 rkqcskedgg gwwynrchaa npngryywgg qytwdmakhg tddgvvwmnw kgswysmrkm
481 smkirpffpq q
```

SEQUENCE LISTING

<110> Vitivity, Inc. et al.

<120> DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE

<130> MMI-004PC

<150> 10/017,724
<151> 2001-12-14

<150> US 60/317,178
<151> 2001-09-05

<150> US 60/329,958
<151> 2001-10-16

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 5784
<212> DNA
<213> Homo Sapiens

<400> 1
acggcatcca gtacagaggg gctggacttg gacccctgca gcagccctgc acaggagaag 60
cggcatataa agccgcgctg cccgggagcc gctcggccac gtccaccgga gcatcctgca 120
ctgcagggcc ggtctctcgc tccagcagag cctgcgcctt tctgactcgg tccggaacac 180
tgaaaccagt catcactgca tctttttggc aaaccaggag ctcagctgca ggaggcagga 240
tggtctggag gctggtcctg ctggctctgt gggtgtggcc cagcacgcaa gctggtcacc 300
aggacaaaga cacgaccttc gacctttca gtatcagcaa catcaaccgc aagaccattg 360
gcgccaagca gttccgcggg cccgaccccg gcgtgccggc ttaccgcttc gtgcgctttg 420
actacatccc accggtgaac gcagatgacc tcagcaagat caccaagatc atgcggcaga 480
aggagggctt cttcctcacg gcccagctca agcaggacgg caagtccagg ggcacgctgt 540
tggctctgga gggccccggt ctctcccaga ggcagttcga gatcgtctcc aacggccccg 600
cggacacgct ggatctcacc tactggattg acggcacccg gcatgtggtc tccctggagg 660
acgtcggcct ggctgactcg cagtggaaga acgtcaccgt gcaggtggct ggcgagacct 720
acagcttgca cgtgggctgc gacctcatag gaccagttgc tctggacgag cccttctacg 780
agcacctgca ggcggaaaag agccggatgt acgtggccaa aggctctgcc agagagagtc 840
acttcagggg tttgcttcag aacgtccacc tagtgtttga aaactctgtg gaagatattc 900
taagcaagaa gggttgccag caaggccagg gagctgagat caacgccatc agtgagaaca 960
cagagacgct gcgcctgggt ccgcatgtca ccaccgagta cgtgggcccc agctcggaga 1020
ggaggcccga ggtgtgcgaa cgctcgtgcg aggagctggg aaacatggtc caggagctct 1080
cggggctcca cgtcctcgtg aaccagctca gcgagaacct caagagagtg tcgaatgata 1140
accagtttct ctgggagctc attggtggcc ctcctaagac aaggaacatg tcagcttgct 1200
ggcaggatgg ccggttcttt gcggaaaatg aaacgtgggt ggtggacagc tgcaccacgt 1260
gtacctgcaa gaaatttaaa accatttgcc accaaatcac ctgcccgcct gcaacctgcg 1320
ccagtccatc ctttgtggaa ggcgaatgct gcccttcctg cctccactcg gtggacggtg 1380
aggagggctg gtctccgtgg gcagagtgga cccagtgctc cgtgacgtgt ggctctggga 1440
cccagcagag aggccggtcc tgtgacgtca ccagcaacac ctgcttgggg ccctcgatcc 1500
agacacgggc ttgcagtctg agcaagtgtg acacccgcat ccggcaggac ggcggctgga 1560
gccactggtc accttggtct tcatgctctg tgacctgtgg agttggcaat atcacacgca 1620
tccgtctctg caactcccca gtgccccaga tggggggcaa gaattgcaaa gggagtggcc 1680
gggagaccaa agcctgccag ggcgccccat gcccaatcga tggccgctgg agcccctggt 1740
ccccgtggtc ggcctgcact gtcacctgtg ccggtgggat ccgggagcgc acccgggtct 1800
gcaacagccc tgagcctcag tacggaggga aggcctgcgt gggggatgtg caggagcgtc 1860
agatgtgcaa caagaggagc tgccccgtgg atggctgttt atccaacccc tgcttcccgg 1920
gagcccagtg cagcagcttc cccgatgggt cctggtcatg cggcttctgc cctgtgggct 1980
tcttgggcaa tggcacccac tgtgaggacc tggacgagtg tgccctggtc cccgacatct 2040

gcttctccac cagcaaggtg cctcgctgtg tcaacactca gcctggcttc cactgcctgc 2100
cctgcccgcc ccgatacaga gggaaccagc ccgtcggggt cggcctggaa gcagccaaga 2160
cggaaaagca agtgtgtgag cccgaaaacc catgcaagga caagacacac aactgccaca 2220
agcacgcgga gtgcatctac ctgggtcact tcagcgaccc catgtacaag tgcgagtgcc 2280
agacaggcta cgcgggcgac gggctcatct gcggggagga ctcggacctg gacggctggc 2340
ccaacctcaa tctggtctgc gccaccaacg ccacctacca ctgcatcaag gataactgcc 2400
cccatctgcc aaattctggg caggaagact ttgacaagga cgggattggc gatgcctgtg 2460
atgatgacga tgacaatgac ggtgtgaccg atgagaagga caactgccag ctcctcttca 2520
atccccgcca ggctgactat gacaaggatg aggttgggga ccgctgtgac aactgccctt 2580
acgtgcacaa ccctgcccag atcgacacag acaacaatgg agagggtgac gcctgctccg 2640
tggacattga tggggacgat gtcttcaatg aacgagacaa ttgtccctac gtctacaaca 2700
ctgaccagag ggacacggat ggtgacggtg tgggggatca ctgtgacaac tgccccctgg 2760
tgcacaaccc tgaccagacc gacgtggaca atgaccttgt tggggaccag tgtgacaaca 2820
acgaggacat agatgacgac ggccaccaga acaaccagga caactgcccc tacatctcca 2880
acgccaacca ggctgaccat gacagagacg gccagggcga cgcctgtgac cctgatgatg 2940
acaacgatgg cgtccccgat gacagggaca actgccggct tgtgttcaac ccagaccagg 3000
aggacttgga cggtgatgga cggggtgata tttgtaaaga tgattttgac aatgacaaca 3060
tcccagatat tgatgatgtg tgtcctgaaa acaatgccat cagtgagaca gacttcagga 3120
acttccagat ggtccccttg gatcccaaag ggaccaccca aattgatccc aactgggtca 3180
ttcgccatca aggcaaggag ctggttcaga cagccaactc ggaccccggc atcgctgtag 3240
gttttgacga gtttgggtct gtggacttca gtggcacatt ctacgtaaac actgaccggg 3300
acgacgacta tgctggcttc gtctttggtt accagtcaag cagccgcttc tatgtggtga 3360
tgtggaagca ggtgacgcag acctactggg aggaccagcc cacgcgggcc tatggctact 3420
ccggcgtgtc cctcaaggtg gtgaactcca ccacggggac gggcgagcac ctgaggaacg 3480
cgctgtggca cacggggaac acgccggggc aggtgcgaac cttatggcac gaccccagga 3540
acattggctg gaaggactac acggcctata ggtggcacct gactcacagg cccaagaccg 3600
gctacatcag agtcttagtg catgaaggaa aacaggtcat ggcagactca ggacctatct 3660
atgaccaaac ctacgctggc gggcggctgg gtctatttgt cttctctcaa gaaatggtct 3720
atttctcaga cctcaagtac gaatgcagag atatttaaac aagatttgct gcatttccgg 3780
caatgccctg tgcatgccat ggtccctaga cacctcagtt cattgtggtc cttgcggctt 3840
ctctctctag cagcacctcc tgtcccttga ccttaactct gatggttctt cacctcctgc 3900
cagcaacccc aaacccaagt gccttcagag gataaatatc aatggaactc agagatgaac 3960
atctaaccca ctagaggaaa ccagtttggt gatatatgag actttatgtg gagtgaaaat 4020
tgggcatgcc attacattgc tttttcttgt ttgtttaaaa agaatgacgt ttacatataa 4080
aatgtaatta cttattgtat ttatgtgtat atggagttga agggaatact gtgcataagc 4140
cattatgata aattaagcat gaaaaatatt gctgaactac ttttggtgct taaagttgtc 4200
actattcttg aattagagtt gctctacaat gacacacaaa tcccgctaaa taaattataa 4260
acaagggtca attcaaattt gaagtaatgt tttagtaagg agagattaga agacaacagg 4320
catagcaaat gacataagct accgattaac taatcggaac atgtaaaaca gttacaaaaa 4380
taaacgaact ctcctcttgt cctacaatga aagccctcat gtgcagtaga gatgcagttt 4440
catcaaagaa caaacatcct tgcaaatggg tgtgacgcgg ttccagatgt ggatttggca 4500
aaacctcatt taagtaaaag gttagcagag caaagtgcgg tgctttagct gctgcttgtg 4560
ccgttgtggc gtcggggagg ctcctgcctg agcttccttc cccagctttg ctgcctgaga 4620
ggaaccagag cagacgcaca ggccggaaaa ggcgcatcta acgcgtatct aggctttggt 4680
aactgcggac aagttgcttt tacctgattt gatgatacat ttcattaagg ttccagttat 4740
aaatattttg ttaatattta ttaagtgact atagaatgca actccattta ccagtaactt 4800
attttaaata tgcctagtaa cacatatgta gtataatttc tagaaacaaa catctaataa 4860
gtatataatc ctgtgaaaat atgaggcttg ataatattag gttgtcacga tgaagcatgc 4920
tagaagctgt aacagaatac atagagaata atgaggagtt tatgatggaa ccttaatata 4980
taatgttgcc agcgatttta gttcaatatt tgttactgtt atctatctgc tgtatatgga 5040
attcttttaa ttcaaacgct gaaaacgaat cagcatttag tcttgccagg cacacccaat 5100
aatcagtcat gtgtaatatg cacaagtttg tttttgtttt tgtttttttt gttggttggt 5160
ttttttgctt taagttgcat gatctttctg caggaaatag tcactcatcc cactccacat 5220
aaggggttta gtaagagaag tctgtctgtc tgatgatgga tagggggcaa atctttttcc 5280
cctttctgtt aatagtcatc acatttctat gccaaacagg aacgatccat aactttagtc 5340
ttaatgtaca cattgcattt tgataaaatt aattttgttg tttcctttga ggttgatcgt 5400
tgtgttgttt tgctgcactt tttacttttt tgcgtgtgga gctgtattcc cgagacaacg 5460
aagcgttggg atacttcatt aaatgtagcg actgtcaaca gcgtgcaggt tttctgtttc 5520
tgtgttgtgg ggtcaaccgt acaatggtgt gggaatgacg atgatgtgaa tatttagaat 5580
gtaccatatt ttttgtaaat tatttatgtt tttctaaaca aatttatcgt ataggttgat 5640
gaaacgtcat gtgttttgcc aaagactgta aatatttatt tatgtgttca catggtcaaa 5700

atttcaccac tgaaaccctg cacttagcta gaacctcatt tttaaagatt aacaacagga 5760
aataaattgt aaaaaaggtt ttct 5784

<210> 2
<211> 1172
<212> PRT
<213> Homo Sapiens

<400> 2
Met Val Trp Arg Leu Val Leu Leu Ala Leu Trp Val Trp Pro Ser Thr
1 5 10 15
Gln Ala Gly His Gln Asp Lys Asp Thr Thr Phe Asp Leu Phe Ser Ile
20 25 30
Ser Asn Ile Asn Arg Lys Thr Ile Gly Ala Lys Gln Phe Arg Gly Pro
35 40 45
Asp Pro Gly Val Pro Ala Tyr Arg Phe Val Arg Phe Asp Tyr Ile Pro
50 55 60
Pro Val Asn Ala Asp Asp Leu Ser Lys Ile Thr Lys Ile Met Arg Gln
65 70 75 80
Lys Glu Gly Phe Phe Leu Thr Ala Gln Leu Lys Gln Asp Gly Lys Ser
85 90 95
Arg Gly Thr Leu Leu Ala Leu Glu Gly Pro Gly Leu Ser Gln Arg Gln
100 105 110
Phe Glu Ile Val Ser Asn Gly Pro Ala Asp Thr Leu Asp Leu Thr Tyr
115 120 125
Trp Ile Asp Gly Thr Arg His Val Val Ser Leu Glu Asp Val Gly Leu
130 135 140
Ala Asp Ser Gln Trp Lys Asn Val Thr Val Gln Val Ala Gly Glu Thr
145 150 155 160
Tyr Ser Leu His Val Gly Cys Asp Leu Ile Gly Pro Val Ala Leu Asp
165 170 175
Glu Pro Phe Tyr Glu His Leu Gln Ala Glu Lys Ser Arg Met Tyr Val
180 185 190
Ala Lys Gly Ser Ala Arg Glu Ser His Phe Arg Gly Leu Leu Gln Asn
195 200 205
Val His Leu Val Phe Glu Asn Ser Val Glu Asp Ile Leu Ser Lys Lys
210 215 220
Gly Cys Gln Gln Gly Gln Gly Ala Glu Ile Asn Ala Ile Ser Glu Asn
225 230 235 240
Thr Glu Thr Leu Arg Leu Gly Pro His Val Thr Thr Glu Tyr Val Gly
245 250 255
Pro Ser Ser Glu Arg Arg Pro Glu Val Cys Glu Arg Ser Cys Glu Glu
260 265 270
Leu Gly Asn Met Val Gln Glu Leu Ser Gly Leu His Val Leu Val Asn
275 280 285
Gln Leu Ser Glu Asn Leu Lys Arg Val Ser Asn Asp Asn Gln Phe Leu
290 295 300
Trp Glu Leu Ile Gly Gly Pro Pro Lys Thr Arg Asn Met Ser Ala Cys
305 310 315 320
Trp Gln Asp Gly Arg Phe Phe Ala Glu Asn Glu Thr Trp Val Val Asp
325 330 335
Ser Cys Thr Thr Cys Thr Cys Lys Lys Phe Lys Thr Ile Cys His Gln
340 345 350
Ile Thr Cys Pro Pro Ala Thr Cys Ala Ser Pro Ser Phe Val Glu Gly
355 360 365
Glu Cys Cys Pro Ser Cys Leu His Ser Val Asp Gly Glu Glu Gly Trp
370 375 380
Ser Pro Trp Ala Glu Trp Thr Gln Cys Ser Val Thr Cys Gly Ser Gly
385 390 395 400
Thr Gln Gln Arg Gly Arg Ser Cys Asp Val Thr Ser Asn Thr Cys Leu
405 410 415

```
Gly Pro Ser Ile Gln Thr Arg Ala Cys Ser Leu Ser Lys Cys Asp Thr
            420                 425                 430
Arg Ile Arg Gln Asp Gly Gly Trp Ser His Trp Ser Pro Trp Ser Ser
        435                 440                 445
Cys Ser Val Thr Cys Gly Val Gly Asn Ile Thr Arg Ile Arg Leu Cys
    450                 455                 460
Asn Ser Pro Val Pro Gln Met Gly Gly Lys Asn Cys Lys Gly Ser Gly
465                 470                 475                 480
Arg Glu Thr Lys Ala Cys Gln Gly Ala Pro Cys Pro Ile Asp Gly Arg
                485                 490                 495
Trp Ser Pro Trp Ser Pro Trp Ser Ala Cys Thr Val Thr Cys Ala Gly
            500                 505                 510
Gly Ile Arg Glu Arg Thr Arg Val Cys Asn Ser Pro Glu Pro Gln Tyr
        515                 520                 525
Gly Gly Lys Ala Cys Val Gly Asp Val Gln Glu Arg Gln Met Cys Asn
    530                 535                 540
Lys Arg Ser Cys Pro Val Asp Gly Cys Leu Ser Asn Pro Cys Phe Pro
545                 550                 555                 560
Gly Ala Gln Cys Ser Ser Phe Pro Asp Gly Ser Trp Ser Cys Gly Phe
                565                 570                 575
Cys Pro Val Gly Phe Leu Gly Asn Gly Thr His Cys Glu Asp Leu Asp
            580                 585                 590
Glu Cys Ala Leu Val Pro Asp Ile Cys Phe Ser Thr Ser Lys Val Pro
        595                 600                 605
Arg Cys Val Asn Thr Gln Pro Gly Phe His Cys Leu Pro Cys Pro Pro
    610                 615                 620
Arg Tyr Arg Gly Asn Gln Pro Val Gly Val Gly Leu Glu Ala Ala Lys
625                 630                 635                 640
Thr Glu Lys Gln Val Cys Glu Pro Glu Asn Pro Cys Lys Asp Lys Thr
                645                 650                 655
His Asn Cys His Lys His Ala Glu Cys Ile Tyr Leu Gly His Phe Ser
            660                 665                 670
Asp Pro Met Tyr Lys Cys Glu Cys Gln Thr Gly Tyr Ala Gly Asp Gly
        675                 680                 685
Leu Ile Cys Gly Glu Asp Ser Asp Leu Asp Gly Trp Pro Asn Leu Asn
    690                 695                 700
Leu Val Cys Ala Thr Asn Ala Thr Tyr His Cys Ile Lys Asp Asn Cys
705                 710                 715                 720
Pro His Leu Pro Asn Ser Gly Gln Glu Asp Phe Asp Lys Asp Gly Ile
                725                 730                 735
Gly Asp Ala Cys Asp Asp Asp Asp Asp Asn Asp Gly Val Thr Asp Glu
            740                 745                 750
Lys Asp Asn Cys Gln Leu Leu Phe Asn Pro Arg Gln Ala Asp Tyr Asp
        755                 760                 765
Lys Asp Glu Val Gly Asp Arg Cys Asp Asn Cys Pro Tyr Val His Asn
    770                 775                 780
Pro Ala Gln Ile Asp Thr Asp Asn Asn Gly Glu Gly Asp Ala Cys Ser
785                 790                 795                 800
Val Asp Ile Asp Gly Asp Asp Val Phe Asn Glu Arg Asp Asn Cys Pro
                805                 810                 815
Tyr Val Tyr Asn Thr Asp Gln Arg Asp Thr Asp Gly Asp Gly Val Gly
            820                 825                 830
Asp His Cys Asp Asn Cys Pro Leu Val His Asn Pro Asp Gln Thr Asp
        835                 840                 845
Val Asp Asn Asp Leu Val Gly Asp Gln Cys Asp Asn Asn Glu Asp Ile
    850                 855                 860
Asp Asp Asp Gly His Gln Asn Asn Gln Asp Asn Cys Pro Tyr Ile Ser
865                 870                 875                 880
Asn Ala Asn Gln Ala Asp His Asp Arg Asp Gly Gln Gly Asp Ala Cys
                885                 890                 895
Asp Pro Asp Asp Asp Asn Asp Gly Val Pro Asp Asp Arg Asp Asn Cys
```

```
          900                 905                 910
Arg Leu Val Phe Asn Pro Asp Gln Glu Asp Leu Asp Gly Asp Gly Arg
       915                 920                 925
Gly Asp Ile Cys Lys Asp Asp Phe Asp Asn Asp Asn Ile Pro Asp Ile
    930                 935                 940
Asp Asp Val Cys Pro Glu Asn Asn Ala Ile Ser Glu Thr Asp Phe Arg
945                 950                 955                 960
Asn Phe Gln Met Val Pro Leu Asp Pro Lys Gly Thr Thr Gln Ile Asp
                965                 970                 975
Pro Asn Trp Val Ile Arg His Gln Gly Lys Glu Leu Val Gln Thr Ala
            980                 985                 990
Asn Ser Asp Pro Gly Ile Ala Val Gly Phe Asp Glu Phe Gly Ser Val
        995                 1000                1005
Asp Phe Ser Gly Thr Phe Tyr Val Asn Thr Asp Arg Asp Asp Asp Tyr
    1010                1015                1020
Ala Gly Phe Val Phe Gly Tyr Gln Ser Ser Ser Arg Phe Tyr Val Val
1025                1030                1035                1040
Met Trp Lys Gln Val Thr Gln Thr Tyr Trp Glu Asp Gln Pro Thr Arg
                1045                1050                1055
Ala Tyr Gly Tyr Ser Gly Val Ser Leu Lys Val Val Asn Ser Thr Thr
            1060                1065                1070
Gly Thr Gly Glu His Leu Arg Asn Ala Leu Trp His Thr Gly Asn Thr
        1075                1080                1085
Pro Gly Gln Val Arg Thr Leu Trp His Asp Pro Arg Asn Ile Gly Trp
    1090                1095                1100
Lys Asp Tyr Thr Ala Tyr Arg Trp His Leu Thr His Arg Pro Lys Thr
1105                1110                1115                1120
Gly Tyr Ile Arg Val Leu Val His Glu Gly Lys Gln Val Met Ala Asp
                1125                1130                1135
Ser Gly Pro Ile Tyr Asp Gln Thr Tyr Ala Gly Gly Arg Leu Gly Leu
            1140                1145                1150
Phe Val Phe Ser Gln Glu Met Val Tyr Phe Ser Asp Leu Lys Tyr Glu
        1155                1160                1165
Cys Arg Asp Ile
    1170
```

<210> 3
<211> 98829
<212> DNA
<213> Homo Sapiens

<220>
<221> misc_feature
<222> (1)..(98829)
<223> n = A,T,C or G

<400> 3

```
actttcttgc tgctctgcct gcatggacct gtgacaggca tcatctatng aatctcatgn 60
agcgtcgact ccctgcccca ggcattggtg agtggcaagt gagagctgct cacaggcatc 120
aaagggtcaa aatagcaccc aaggccgggc atgannncac gcctgtaatc ccagtacttt 180
gggaggctga ggtaggcgga tcacctgagg tcaggagctc aaagccagcc tggccaacat 240
ggagaaaccc cgtctctact aaaaatacaa aaattagctg ggtgtggtgg tgtgtgcctg 300
taatcccagc tacttggagg ctgaggcagg agaataactt gaacccggga ggcaaaagtt 360
gcagtgagca agattgcacc attgcactcg agcctgggga aaaagagcaa gactccatct 420
caaaaaaaaa aaaaaaaaat tgcacctaag cccaagccca gaaggtggcc ccgaacctgg 480
gccttccttt gaagccaggc ctggtatttg ccaagatcgt gcctcctgcc cccacatcac 540
caacgttaac tgccctccat tagctggtgc tgcctgggtg ctgcgtgggc gctggcgtgt 600
gaaatggcaa catgaggccg tgtctactga gccaggcatc ttagtgcttt acacagagaa 660
gtgaaacctc ctagcaaccc tgtgagttag gtaccatttg attcccattt ttcagataag 720
gaaaccaagg cacaaagtgg ttggtaggtg gactaaggtc acccaacaaa tggaggcaaa 780
```

acgggacacg tgtccccacc aggagccacc agctgtccct ccttctccat cccacagact 840
gcgtgaagcc ctgggctact ctggctggat ggtatcaagt gtcacattct gtggagccag 900
gatgttcctc caatccttta atcccaaatc ttttattgtg tttgttttgt tttgttttga 960
cagagtcttg ctctgtcgtc caggctggag tgctgtggtg caatctcagc tcactgcaac 1020
ctctgcctcc tgggttcaag tgattctcct gcctcatcct ccctagtagc tgggactaca 1080
ggtgcgcacc accaagcccg gctaattttt gtatttttg ttgagacggg gtttcaccat 1140
gttggccagg ctggttatga actcctgacc tcaaatgatc tgcctgcctc agcctcccaa 1200
agtgctggca ttacaggcat gagccaccat gcctggccta atcccaaatc ttaaacctac 1260
ttttcctggt tctgcctcta gtgaaatcaa agatccctcc gctgctgtcc ctatgggaat 1320
gtggacaggg atgacacctc tcgccagctc tgggtggccc ctctggtctg gtcctttgct 1380
gcccagctcc tcaagtcctc ctcaacaccc ctcttctcca gccctgctgc cactctcccc 1440
acctgagtac agaccctcat catttcccac tgggcccccc cttctcgtcc ccccagtccc 1500
tnctccctag ntccatttat ttattctnat tttgnnagac ggagttttgc tcttgttgcc 1560
caagctggag tgcaatggcg cgatctcagc tcactgcaac ctctgcctcc caggttcaag 1620
cgattttccn tgcctcagcc tcccaagtag ctgggattac aggcatgcac caccacaccc 1680
ggctaatttt ngtattttta gtagaggtgg ggtttctcca tgttggtcan ggctggtctc 1740
gaactcccga cctcaggtga tccacctgcc tcagcctgcc aaagtgctgg gattnacagg 1800
cgtgancanc tgntgnccct gcccctagtc catttttttt ttttttaatt gatcattctt 1860
gggtgtttct cgcagagggg gatttggcag ggtcacagga caatagtgga gggaaggtca 1920
gcagataaac aagtgaacaa aggtctctgg ttttcctagg cagaggaccc tgtggccttc 1980
cgcagtgttt gtgtccctgg gtacttgaga ttagggagtg gtgatgactc ttaaggagca 2040
tgctgccttc aagcatctgt ttaacaaagc acatcttgca ccgcccttaa tccattcaac 2100
cctgagtgga tacagcacat gtttcagaga gcacagggtt gggggtaagg tcacagatca 2160
acaggatccc aaggcagaag aatttttctt agtacagaac aaaatgaaaa gtctcccacg 2220
tctacctctt tctacacaga cacggcaacc atccgatttc tcaatctttt ccccaccctt 2280
ccccctttc tattccacaa aaccgccatt gtcatcatgg cccgttctca atgagctgtt 2340
gggtacacct cccagaccgg gtggtggccg ggcagagggg ctcctcactt cccagtaggg 2400
gcggccgggc agaggcgccc ctcancctcc cggacggggc ggctggccgg gcgggggggct 2460
gacccccccа cctccctccc ggacggggcg gccggccagg cagaggggct cctcacttcc 2520
cagtaggggc ggccgggcag aggcgccccт cacctcccgg acagggcggc tggccgggca 2580
gggggctgat ccccccacct ccctcccgga cggggaggct ggccgggcgg ggggctgacc 2640
ccccctgacc cccccacctc cctcccggac ggggcggctg gctgggcggg gggctgaccc 2700
ccacacctcc ctcccggacg gggcagctgg ccgggcgggg ggctgacccc cccacctccc 2760
tcctggacgg agcggctggc cgggcagagg ggctcctcac ttcccagtag gggcggccgg 2820
gcagaggcgc ccctcacctc ccggacgggg cggctggccg ggncaggggg ctgatcgccc 2880
cacnctccct cccggacggg gaggctggcc gggcgggggg ctgacccccc cacctccctg 2940
ccggacgagg tggctgccgg gcagagacgc tcctcacttc ccagacgggg tggcttctgg 3000
acggatgggc tcctcacttc tcagacgggg cggttgccgg gcggagggtc tcctcacttc 3060
tcagaggggg cggccgggca gagacgctcc tcacatcccg gacggggcgg cagggcagag 3120
gtgctcccca catctcagac gatgggcggc cgggcagaga cgctcctcac ttcccagatg 3180
ggatggctgc agggaagagg cgctcctcac ttcctagatg ggatggcggc caggcagaga 3240
cgctcctcac gtcccagacg atgggcggct gggcagagac gctcctcact tcccagacgg 3300
ggtggcggcc gggcagaggc tgcaatctcg gcactttggg aggccaaggc aggctgctgg 3360
gaggtgaagg ttgtagcgag ccgagatcac gccactgcac tccagcctgg gcaccattga 3420
gcacggagtg aacgagactc cgtctgcaat cccggcacct caggatgccg aggctggcgg 3480
atcactcgcg gttaggagct ggagaccagc ccggccaaca cagcaaaacc ccgtctccac 3540
caaaaaaata cgaaaaccag tcaggtgtgg cggcgcgcac ctgcaatcgc aggcagtcgg 3600
caggctgagg caggagaatc aggcagcagt accgtccagc ttcagctcgg catcagaggg 3660
angaccgtgg agagagggag agggagaccg tggggagagg gagagggaga gggagggaga 3720
gggagaggga gagggagagg gagagagcta gtccattttt atanncatgg ctctgaggat 3780
gctccaatct gaccacagct tctgttcact tagcgccctg ccccggtgtc tcatcacctt 3840
ggggataaaa tccacactct gagagtgagt gcaaagtcct tcctggcgtg gcctctgcct 3900
cctccatttc ctcaggcctt gctctaccac acaagtccct ctattagtgc ctcccgatgg 3960
ccagctcccт ggcagggctg atagtgataa aatgttcagt agtcagcaat aaagaggcaa 4020
aaacagcaac actgatttat agatatgaaa tcaaccagat gtatcagcca agttgccagc 4080
attcctttct tggtcacaat atcttttatt ccaaacatcc tttctattta ttttttgtag 4140
ttaagatttt attacttctt ttgtaaaacg gtggtgagaa taagggacta tttttttaa 4200
atgtccttac ttggcaaact aaaaaatgtt tgcaactcta gtaagactcc actgggagct 4260
caggggctgg ggtgcctgat ctggccctaa agccccannn nnnnnnnnnn nnnnncngac 4320
actgacctgg tgcctgcgtg attcttcttc attgttccca gacaggggcc aggaaatgga 4380
atgaaagcag ccactgtctg aagagctgga gaccatcatc tgcctctgga agcccagaga 4440

acctcggctc agacagaagg acagagactg agggaaggga gagagactgt gacagagaag 4500
cagaggaggg tgacagagtc agggaggaac aaaacagcct gcagtgggag cagagacaga 4560
aatgtggggg acccacaggg angggggaggg agggaagggg agggacggag ggagggacaa 4620
ctgccgtcca agtggctgtg agagcctggg gctggggaga ggcaccctcc tcctgttggc 4680
ttctcataca gtctctatca ggggacccag gacacaagaa gcaaattcat ctttgggtcc 4740
acccttcagc tctagaagtc tagggcctgg ccaggtcctc tgaaggggtc tctggcccca 4800
ggcagcttcc tgcccccctg cgtggccccc caagcccttc tcaagagccc ttcagtaaat 4860
aaattcactg gaagacatca atgatgggtc ataggaggtg ttcctagggg agaaaatgac 4920
atttagtttt tagctcactt ggagccattt gctttaagaa gcacctcttg gccaggcgca 4980
gtgactcacg cctgtaatcc caacagtttg ggaggccgag gtgggcggat cacctgaggt 5040
caggagttca agaccagcct ggtcaacatg gtgaaaccct gtnctctact aaaaatacaa 5100
aaattagttg ggtgtggtgg catgtgcctg taatcccagc tgcttgggag gctgaggtga 5160
gagaatcgct tgaactcagg aggtggaggc tgcagtgagc tgagatcaca ccactgcatt 5220
ccagcctggg tgacacagtg agactatgtc taaaaaacaa aaaacaaaaa cgaagcacct 5280
cttgagagag ctttgcatgc agatgtcaca ggcaaaatgc ctcatgttgt ggggaatgtg 5340
gcagtggact tctggggcca ggctgaggcc cccacgtgct gtccccctg ggagcagggc 5400
agggtnggct gggttgacag atgtgcagag aaggagagaa gtgggggcaa tgccttctct 5460
ctgatggtca gctggcggga aaggctccag gggcttctga ggagagaccc taagaggact 5520
ctggatgggc ctcggctggg ggnaggctgg tgctgacctg gcacgtgggc actgtggggg 5580
tgcaatgggg gacagtgtgg gaggtgtcat gggggggtctt tacactgggt actaaagagc 5640
ttaggagcct gagctacctg gtgtggggca atcacatttg ctgctgtgtc caggcaagga 5700
ggagcaggtg ggcttggacc ccaaagtgac atgaagtggc caggactgag cgctgggcca 5760
acgtgagccc caggggggtgc ttggatttat cagctgaaga agggcctatg gaaggggaga 5820
aaacaatcct ccagacagat gagagcactc agtcacatcc acctggacct gacatcacct 5880
cccaggctgt gcaggggcac acacagaact ggaatgggta atgacttatt agagcctcct 5940
tccaacagct gtgggccctg gaacctaaga acccaccttt ggagaacatt tttgggtctg 6000
agagctttga ggccaggcct tgagacaaac ccacccagaa gccctgagct acttttgctg 6060
taagagtagg nagggccctg cccctgctcc aggagccgcc ccttgatccc tggaagagaa 6120
ggaagcactg actgatccag aggggcgatg agtccagagg ctgagatcag ctggaaccgg 6180
gccaccagct gagcgtgtga ttcctctggc tccaggccac ctctgagggt ctcccaatgg 6240
ctgggaggga ttagaggact gtggcttctt gggtgaggga gcctctgggt tacagtcacc 6300
gtgcccttg aatctaatct gctctcgaca agaatcagct gacctgatag gcatgctcag 6360
gcaggaggca natggccagg gccaagagaa gcataaggat ggagtagagg agggagcaga 6420
gcagggtcag gtgaaatctc tcactcagga caattgtgca gctgacttgc tgataccttg 6480
gaggagtcac gaagttcagt gtgcccagtt ggggcatctt ccaggctgtg gctccatcac 6540
cctgcagttt tggctcagaa cagccaccac atcctcactg ctctcctggg atgagaaggt 6600
gggtggagaa ccaggggctt ggtgcgtggc cagagctgga gcagcttcag gagccaggtg 6660
aaacctcaga gcctcacagg ctttgaacaa agtgcaggag tcccagccct gccatttatg 6720
gcgggttcct ggcggtgatc agtcattatc agcagtatgt tcactctgct gatatcaccc 6780
aaggccatga tcatccctta ccccaggcct tgcacacttg cttccctgtc cnnnnntaga 6840
cagctgggnn nngctgttcc tgctcttcgc tgcccctggc actcacagag cagctgactg 6900
actgacctac caacttaggg agaaaagtgg ggaggtgggg atggaggaca ctctggggat 6960
acagcctctt cgtgatttcc ttcctggagg ggagaggcgg ccagggcaat gcaatgctgg 7020
ggctgcagtg cctccgctca gccacagcac tggctgcagg gcaaaccctc aggaccagat 7080
gaggggagct cagcctgatg tccaggttcg ctggcatgtg agggacgtgg tcaggcacag 7140
ctatagccat gggtctatgg acggttcctg gatcccagga attgctggtg gggtccctga 7200
tacacacctc acgtcttcaa cctctgccct ctctgatggc ctgttggcac tcctcctgcc 7260
tatctctgca cactcagcga ctttcccctt tgatggggac aggagtggag gagggcaggg 7320
caggatggag acaccataca cattcctgct cagtgccttt gccctccctc ttccatccat 7380
gggagtcacc tcccctagac cttcantgtg gcctgcgccc ccacttctcc ctcccagagc 7440
cttcctagtg cccccaacaa gtgcctgccc tcgcctgacg ccctgtttgt ccccttctgc 7500
ctcttcttca tgctgcttgt taggttgaaa cataacatct gtgtttcaaa acacatgcag 7560
tcaggtttct gtgcatgtgt gtatgtagat atagattgtg gtgcaatctc agctcactgc 7620
aacctcctcc tcttgggttc aagcaattct cgtgcttcag cctctcgagt agctaggact 7680
acaggcatgc accaccacgc ctggctaatt tttgtttttc tttttgtatt ttttttagtg 7740
gangatgggt ttcgccantg ttggccaggc tggtctggaa ctcctggcct caagtgatca 7800
gcctgccttg gcttgccaaa gttctgggat tacaggtgtg agccaccgca cccagacgtt 7860
agtttacttt ttaattccct gtctacccac ttcttgtaag ttacatgaaa gcagacacct 7920
tgtgtgtctg atctgttact gtctctccag ggcctaggat aagcaggtgc cagtaagcct 7980
ttgttggaag aaaggaagga gagaaaatgg acctcctctg gatgggatcc ttggctcctg 8040
cgtgctgatg tgcagcctgt gacccgggct ctgcactgct cctctggctc caggcaatca 8100

tcatggaaag ggcggcctcc ccagagggct gcctgggaat gaactgggca tctggggaag 8160
cccattncct ttggctgcct cttgcctgtg gcgcccaggg ctgccgtgcc acctcctttg 8220
tgaagctggt ccccagccct gcttatcagc cccatgtccc attcccttgg gcccagcaga 8280
ggtaaaagtg gctaagtggc ttaactgggc caggtgcagt ggctcacgcc ggtaatccca 8340
gcactttggg aggccaaggc gggtggatca cctgaggtcg ggagttcgag accagcctga 8400
ccaacatgga gaaaccccgt ctctactaaa aatacaaaat tagccaggcg tggtggtgca 8460
ttcctgtaat cccagctact tgggaggctg aggcaggagg atcacttgaa ccacggaggc 8520
ggcggttgca gtgagctgtc attgcactcc agcctggaca aggagagcga aactccgtct 8580
taaaaaaaaa aaaaaggtgg cttaactgtc ctctctgtca tggcctctcc ctccctcttg 8640
ccccagtggt atttggctct tttgtctgtc ttgtttttcc cagactgacc agccttgggg 8700
ttaaccaggg gcagggtgac aactcacagg caatgacgtc cctcagcctc cttcccccca 8760
cactccagag gctgcaggtg actcagggtc ttcgtcggtg gctccggaag ccttgcaggt 8820
gccctgccac atggccgctg tgctgctggc tgctctgacg ggtcctgccg gggctactgc 8880
tttcctgtcc cctcccccga agctgggaga tggatttcct ccctgagtga ttaattggtc 8940
tattttaagc tcatcaaagt agcctcgtac tttcaagttc ctcagggagc aggcccggaa 9000
gatgatacct gcggacttaa ggctgctcca ctgcaggata tggggtcggg ggctgtgtct 9060
gtcaattgtg aggtccctca gggcaggccc cctcctgagc tggacagggc ttggctcctg 9120
agtgctgatt ctgcaggctg tggctcagtt ctgcactatt actctggctc cggcaaatcc 9180
tcatgaaaat aagcctggga gccccgaggg ttgcaggtta gccccctaca gccttccctt 9240
gcatcatcct tgctcctcta accagggaca ctcctgcctc ctccctgact atctctgccg 9300
ggttaacacc tgagcatctt ctggctcgat aattctgtca tgacccaata cgccttccct 9360
gaccctcaca tctgggttga gggccttcct atgccctgta cctaatcctg gttttccctc 9420
tgtaaggtgg tcagagatgg gcttgggcag agtagtttca aaggctctcc ccaaatctct 9480
gtgcagctcc cctgactgaa ggcctgtgag aacagggtgg tgcctgtctt gttcactgtt 9540
ccccaatacc tcacctagtg cctggcacag agaaggtaca aaaataaaac tttcccaaga 9600
atgaatggtg tgcatttgaa tcttatagtg tctgaggact ggggaaccat ctcttatccc 9660
tttgtatcct ctgctgccca aggtcagggc ctctgcctga agtccctact gaatgagggt 9720
taactggttc agagatgagg gaggaagcca ttgctgggag gggctagaan ctgggcccct 9780
gcccttccca cacccgagca agagcagtag gattgtcaga tgaaacacag gacatctgct 9840
cacatgtgag gaatacatct ttttttaag acagattctc tctctgtcac cccaggnctg 9900
gangtgcagt ggcatgatct tggctcattt caacctctac ctcttgggtt catgccattc 9960
tcctgtcgta gcctccccag tagctgggat tacaggcatc caccaccatg cctggctaat 10020
tgttgtgttt ttagtagaga cagggttttg cctcattggc aggctggtct tggactcctg 10080
acctcaagtg atccgctcac ctcagcctcc caanagtgct gggattacag gcatgagcca 10140
tggcacccag ctggaataca tctttagtat actgcatgta atatttggga catgctcatg 10200
ctgaaaaaac taactgaatt tgaatttttt tttggttttc ttttttgag atggagtctc 10260
actccattgc ccaggctgga gtgcagtggc acaatcttgg ctcactgcaa cctccgcctc 10320
ctgggttcaa gcaattctcc tgtctcagcc tcccaagtag ctgggattac aggtgcctgc 10380
caccttgccc agctaatttt tgtattttta gtagagacgg ggtttcacct tgttggtcag 10440
gctggtctcg aactcttgac ctcaggtgat ccacccacct cggcatccca aagtgctggg 10500
attacaggtg tgagccaccg cacccggcct gaatttgaat ttaactggga agccttttt 10560
tatttttcc tcttctttct gaatctggta accctagcaa gggagaagtg actagccagg 10620
tgccaggtgg gtctacggcc tgagtctcca gctctctatt cacattctgc ctgcttcctt 10680
gatgcttcct cctccttaaa tgctttctgt cagagaccag ggggcactct aggcctggct 10740
gcccagcacg cggaatccag gagcactcct acctcctaat ccctgcccca gctgccctct 10800
ggatggtcag ggaaggcctg aggaggtgat gtttgagctg agatttgaga gaatgtgacc 10860
agagatgatc caggaaagaa cattccatgg agaggtggga gtgaacttgg tgcatctgaa 10920
aacagaaagg ctagcgtggt tctgcgatgt gggaagggag tggaaacatg gaaacacagt 10980
cagcatttcc ttttttttt ttttttttt ttgagacgga gttttgctct tgttgtccag 11040
gctggagtgt aatggcgcgg tcttggctca ccacaacctc tgcctcctgg gttcaagtga 11100
ttctcttgcc tcagcctccg gagtagctgg gattacaggc atgtgccacc ccgcccggct 11160
aattttgtat ttttagtaga gatggggttt ctccacgttg gtcaggctgg tctctagctc 11220
ccgatctcag gtgatccgcc cgcctcggcc tcccaaagtg ctgggattac aggcgttagc 11280
acnaactcag gaggccctgg aactcatctt gtgccactca agaaagggca ggaggggtcc 11340
tctcctttcc caccttccc aagctgaagg actgcacact tctcaggctt ccaaggcagg 11400
agcttcggca gcaggaacag ggttgcaagg tagttagtct gggtgtctgt cgtcctgggt 11460
tgggacatct gggagctggg gcaaggcctc tgggccctag tgccagagag cagtttggtg 11520
gcgttgcctg ttctttaaaa tgggttaggt gtagccatga gccccccccat gcagcgggtg 11580
tgcaggcctc tccctcagtc agaacaccct cacccacccc tccccacttg aagcaggccc 11640
ttgagaacac tgctggcagg gaggcggctg ctcttcctcc taagggaaaa gcctgcagcc 11700
ccgtgctgct gggccgtggg gtgaggggga agccaaaggg cccctatcac ccctcagcag 11760

gagtctctcc tggctttgga aggggctgaa ggcccgggng ctctcacgag ggaactcagc 11820
cctgctcagc cggtcaggac tcctcctcca cgtctaaatc cacagaccaa acaaaggaca 11880
tggctcagtt ccaccttcaa tccacaggga ttctgcctgt tagcacgcta aggaaactgt 11940
ttcttacttt aaattatcca ttaatttcct ccaactgaaa gatactctac aaatttttctg 12000
ttgttggagg gtttaagctt ggtccctctg agggggtggg ccctgctgtg tcggcccctt 12060
gtgtctgctc gtggaaggtg tgccatgtgc caccgtgtgc catatgcagc tgacacgcct 12120
gtgcgtctcg cctggcgctg gagactgctc tctgtgtact aagcgagtac tggatcccca 12180
gggcctatct tgcccactgc ctgtggcaat nacacaggca atacacaaag cgcattcaag 12240
taagcaggtc tccctgcccc ccggcctccc agaacctcgg tgcctggagg gcaggntggg 12300
ggnagggcgt tcctgaatgg aggccggttt ccctctgcgg ggaggaaacc ccatttccct 12360
cctacccctа ggcgcttgct acatctggcc cggcatcaga caggaaagaa ccctttcttc 12420
ccagcttgta ggaagtctga gctgggcccc ttatctaccc acagcacttc ctggaccggg 12480
ggcctggacc gggctccgaa gggnctggnc tcccggtctt ggtcttggcc tcggcgccct 12540
cgtttctctg cgctttgggc aggggaagct gcggacggca agctctcggc tttcgtgaga 12600
gcctggtgga atcggtgttc cccgaaggct actctgcggg ggcgggggct agntccgggc 12660
cttcccggca gtatcgtcnt cccccgtggg ggctggggna nggtngcccc cagggctnnc 12720
ccggagcggc ggcgtggcca aggcccgaac cgggtctgac atctagtggc ctcctgggcc 12780
cgggcagggc gagggncggg gcagggaaga aggtggaggg cagaatggga ggcgtggagc 12840
gagaaaatca ggaagcgcgg gaccaagccg gggaagggcg gcgggtctcg cccctggcac 12900
ccgctctcct cggggcccgt cccatccccc aggcctggcc gaccccaggt ccttccgtgc 12960
gcagtcgggg ctcgcaagga cgaatcccgc ggccctcgag gcaggcccgg gggagctccc 13020
ggccctccac ccccgcaggc cgcagatccc acgcaccccc gatcatgggg gccccggagg 13080
gaggtcgcga ggccgggctc acggtggccg cggttcgccc acgtgcgggc ccggagctgg 13140
cagggcccgg tcccgagcgt ggccgcaacc gcggggacct ggcacgtgcg gggcagcccc 13200
gggggccgcg tacgccactt ccggtcccgg agacaccgcc cagcccgccg cccggttgcc 13260
atggcgacgc cgtcgcgcca cggcccgcag aaccggccaa gcgacccggg cgcggcgcgg 13320
ggaggctgaa gggacgctcg ggtaggcaag gtaggaggcc gggctggggg tgggagcgga 13380
gcgcgcaggg gtgcgggcgg ggggcggccc aggtgagccc tgactgcgca gggagggaca 13440
gcgcgggggct cccgagtagc agccggcctc gcacctgccc cttgcggccg cgcactggac 13500
tgcggcgccg acccgcaccc tgggcccgag ggctgcaggc tcgcccgccc tctcggagcc 13560
gagcctctcc cggtccaggc ggcccctgcc ctggcctgcc cgggtgccgg gtctctgctg 13620
aagttagaac cgagaccccc gcttgctggt gaccccgagt ttggatcttt gctcctgcgc 13680
cgcgttctag agcaaggtag ggtgcaaatg gaccaaccag cctttgccac ggctgccctc 13740
ttgaaatctg aaaccggagc caccgcctct tgccccgagc agggcccggg ctgcaacgtg 13800
gagccgcagg tccccgcctg tgtctcccga cgcccccagc ttctgagcgc gagggtggga 13860
gtttcccgag tgggaaagcc ccatggcttc ggtggcctcg gtggccctgt ggtgggtcag 13920
gccggtgcca actgcgctga gggcggactg ccgccacctt gacacctggg agatggcagg 13980
gccaccctcc ctccctccct ccctgccctg tccccgactg tatcacggag cgaggatcat 14040
ccgtgtggat ctggggtccc cttcccagac ttctcctttc tggtcgtcct cctctcattc 14100
attccacctt ggcagccttt gtggattttc tgactccgta aatgaagtta catctagggc 14160
gcctagattc tgccccttg gtttaatagc taatgaaaaa aaaaaaaaaa aagctttgcc 14220
acctcctcta ggtgcctgga tttctcccgg tgaaatttga tggaggcagg aaaatacgag 14280
gcggtggcag ccccacctcc ttgcatgagg gagcctccta gagctggagg tggaaccacc 14340
ctgcttgtcc cactcctttt ctttcttgct ttattgcttt cttgctttcc ttactttcct 14400
tccttctttt ttttttgaga tggagtcttg ctctgttgcc aggaggctgg agtgcagagg 14460
gcaatatcag ctcactgcaa cctccatctt ccgggttcaa gcgattctcc tgttccagcc 14520
tccctagtag ctgggattac aggtgtgtgc cacaacacgc agctaatttt tgcatttta 14580
ggagagacgg ggtttcacca tgttggccag gctggtctcg aactcctaac ctcaagtgat 14640
ctgcccgcct tggcatccca aagtgctagg attacagacg tgagccaccg tgcccagcct 14700
tgtcctacta cttttctgcc ctcaccactt ttccttggca ggcagagctt gggctcctgc 14760
taatactggg atggttcagt ggcattagat ggcattggtg gaggggaatg agaaattaac 14820
cccctctgaa agcgtagctt cactgggaaa aagcactcta ctgtctgaac aattcataaa 14880
tgaattttgg gaatatacaa tgtcacagat taaaagcaga ccatcctgtg ttcttccagc 14940
cccaggagct tctgcttggt gtgcaggggc taatggtgga cagcaatgga gtgtagattt 15000
attagagacg gctgccatct ggctctctgt ggccccaaac ctcatgtgtc agcagtgtga 15060
ctctttctaa ggtgctctta ggaccagatt tcgtttttgt aagattggtc tttgatggaa 15120
tgcctacttt gtggcaggca ctgtgctggg cgctggcaca cctttatcca cttactcctt 15180
ctgtgttcaa taaggtagct ccagatatcc tccactttat agattaggaa gcaggcttgg 15240
agaagggact tggcttgtcc aggagctgat ggttataaat aatgagtctg ttttgaaatc 15300
acatctctct aacctgggct gcgctgggcc tttcatgggc cgttggccct ctggccttca 15360
cgggccccgt cttccataaa aaaaataata acaattggcc cggcgcagtg gtgaaagcgc 15420

gtctctacaa aagatacaaa aaattagcca ggcatcgtgg agcacgcctg taatcccaga 15480
ctcggggagt ctgaggcagg agaattgctt ggacccggaa ggcagaggtt gcggtgaccc 15540
gagatcttgc cattgcactc cagcctcaaa aaaaaaaaaa ttaataaaaa ttatatttta 15600
caactgcatt ggtataaata caagtatagt ccagactgga ttacaattaa ttttaaaaac 15660
aaaacatttt agggctggct catgtctgtc tgtaatccca gagctttggg aggctggggc 15720
aggaggactg cttgagacca ggagtttgag actagcctgc acaacatagt gagaccctat 15780
ctctccaaaa aaagaaaaaa ttaactgggc atttcagcat agctgtagtt ccagctactt 15840
gagaggctga ggtgggagga tccctcgagc ccaggagttc gaggctgctg tgagctatga 15900
tcatgccatc gtattccagc ctgggtgaca gaggtagaca ctgttaacaa caacaacaaa 15960
aaaattaatt aaaataagtt ttaaaaaatt gcaacatttt cctcaaccct aaatgttcat 16020
ttttttcttc tgattttcat ggaaatggaa acattttggt ttgtaagcat tgggagccct 16080
ctgtcttgtg ggagaaagga gctcagtgtt tgacttcagg ctctgctgct tcctgggtac 16140
ctagttcctt ggtacctggt ccaagcagcc tggatacccc gggtccctgg gctgcctggg 16200
ccaggacagc cgccctagga taaatggaaa tgcagcccct gcggcttgtt caccctcctg 16260
taatgccttc tctcttaccc tttggtcagg gcctggtgtc tctgcatgca ccagggctta 16320
gtgccatgct ggtacaagcg tggggaggat gcaggaggac tggcatggga tggggagcac 16380
tgcctgggaa gccccttctt cctggggcta ctggggtgct aagaaagtac cctaatagct 16440
cacaccacac tttgctttct ccttcatcac atgctaggac gctctagggc aattgagctg 16500
ttttcccctt cacccatcaa gaagagatat cggcctttcc gcagtcacag tctttcctaa 16560
ctcaattcaa actctgctgt tgataaggag ggctgtagcc agcccaggct gcccgcttcc 16620
cagcctcccc tgctccgcct tccgcccccg aggctgtgca ccatgtgcag tgtctggtcc 16680
ccaataatga gattagtctt ggttgccttt taataaaacg cagtgggcac cgggagggag 16740
agcgatgctt ggctcagtga agatctgcgg gtcatgctgt ccctaatgcg ctgattgcat 16800
taagtggatt ctggctgcag gtagggtgag tgggtgggga cgagggtgac tctcacagct 16860
ctaagatcca gaaactgcca gagatctgtc accctcatcc tgagctgtca cagaggaagg 16920
caagtgactg tgtgaggggc tacgtgagct ccctctggtt gcaaggttct ggtctgcagg 16980
gcagtggagc ccttgggggt ggggagtggc agcttccagg ccttgaagct gctgcccaca 17040
gttctgctct gagaacacag agggcccaag aacagccggt gtcccagggc tgcccagtga 17100
ggagggaaga gcgagagagt attcttgctg tcaagactgg gaaatgaggg ccacgattca 17160
aagccttgct tcctagggag aaatctcacc caatgtccag gtttgcaaat gcagcagaca 17220
catttgtggg tgggtcagat tctgtccaga gataccagaa tgtctatccc tgtacacccc 17280
cacctctggc atgggacagc ccttccctgc acacaaatgg gttatcaatt atgtccaatg 17340
aatggcctca cgagagtcct gttgcggaag ggatcttccg cttatctcct gtaaagcaca 17400
gagccacgct cagaacctac ggactgtggt cgcctgcagg ccagccatgg gctttccttt 17460
catccttaca cttagatttt gaagcctcgc ctttgaatca gaagcaggct gctgtcattc 17520
cagaggccgc tgttcatctc tggcttcctc ccagatgccc aggtgccccg ggtgcctggg 17580
ctgcctaggc cagggcaccc gccttaggag caaatggaaa tgcatcttcc tcggcttgtt 17640
tgccctcctg taacccccttc cctcttgcct tcggtcaggg cctgctgtct tgcacgcacc 17700
agtgcttagt gccatggtgg aaccgaggcg gggaggacgc aggaggactg gcacaggagg 17760
gggggcaccg cccgggagct tcccctgcgt caaagcagct tcctcagtga gctccaacac 17820
agacccttgc tggcctcagt ctctggaaag actgacagca ccaccctagt tcctgggcct 17880
tggaaaggcc aaagccacga gcaggcagcc actgttcggc aggagagcta gagatgctat 17940
tgcccacaga catcaaagca ggggtgccgt cttccgtgtg catgctggga cgggtcttgg 18000
aggcccagat tagctcctgg cttgaacttg taccaccctt acctgaaaat tgggtctgta 18060
ccaaagctca tctgcacgcc cctttagggc agggaccagg ccccctctgt gtggccctag 18120
cacctggcca agtgcaggtg gctgcgggtt gagttgagct gctatttgct gatgggaatt 18180
tgggggcctg tcgcttgtcc tctttgagcc tccatttcct ggcctgtcaa attggggcaa 18240
tgctcctgtt atccacacag tgtcatttaa gggtgacatt agagatggca tgaagtgccc 18300
agcacagagc aggtgctcac aggctgttgg gtccccaccc tactgacaca ccaagcacac 18360
gtctgccctg ccatggtggg gagcaaacaa tgttatggtt ccgactcccc gggaggccag 18420
gccgggcatt tcactgtagg atgttgaagc cgcaggcagt cgtcgctgcg agaagagagc 18480
ctggctctga tgtccactgt cttccaggga gcctggtggg aaggaaggag tgggcagcgg 18540
cccctcgctc tgcgggcctc tcctgccctt tgtactccac gaggtgtgag gaagttgccg 18600
ggtcacccag cagagggaga ggtgatgtcc cctctgttct tcctaatttg tagctaatca 18660
gattctgcct gcagccaggc tctctgggga tgaatctaat aaaggaggcg gcgacctgga 18720
gaggaggaca gacaacaagt gaccaggggg gacagcactt ctttcagtac attttcctgt 18780
tagactcggg gtattagaga cccaaagata cccaaggagg agtgagagcg tccctgcctt 18840
cgaggagcag acagggcacc gcgagggtaa ggaggctccg ggagagcctg gggacagagt 18900
gatctgccac cgcggtgcct gaggagtgga ggcaggggtg tgggcccaca gcagggagcc 18960
aggctgggcc atgtccttgg attctgtgag cccaggcgtc tcctctagag ggcctagcag 19020
gaacagtggg ggctgcagat ggaaacgggc aggatcatgc catgcagttg gcatttcact 19080

ggcactcggg ggcacttacg caggggacag catgatgttg ggagctggct ttgtccctgg 19140
catgtctgtc acgttgaaga gggagaagct gcaggccctg acttgggtgg ggcagtgagg 19200
atgctcagga gggaagtgag gggaggggtg ccagcatgcg gcggcctggg cagccttctg 19260
tggaggagct gaggtgtcac ctgggcccgg aagaagggca gccctgctag ctagtgggga 19320
gggcatgaga gaacttggcg acgatgggag agaaggtgac ttgggacgag ctgggtttga 19380
agcgtctgtg ggatgtccag gtagagatga gtagtgaacc gtcagaaatg gggcttgaat 19440
ttggggtggc ggggacagaa ctgatacccg tgtaggtgac agataaggac agaggcagga 19500
gttccaaaaa taggatttat tgttggctgg ggcttctgga atcccgtatt cccgtccacc 19560
ctgggaatgt ggattggagc tatgtttctg ggagatgatt tgatggaatg tgtcagaagt 19620
ctaaaacaga acctagccct tgtttcagga ttccctgcag agaagtttgc agatgaccac 19680
atcctgttgt ttatttggtg gaaaactgaa atcgaaagct agctaaatgc tgaacaatgg 19740
gaactagttg gctaaaaaat ggccgtgcaa tgcaagagtg tgaggttacg aaaatgatga 19800
tgtgggcgcg tatccattga cacaaacatc catgtgtgct acgccctcag ttaaaaagca 19860
ggttagaaaa tggaataaat gatcctgctt ctataaaagg atgcatatgt gttttaggaa 19920
gaaggtgtgt gaggatatag atcaaattct gatggcctgt cattttacct tttttgctga 19980
tccatattct actgcagtgc gtgtggatta ttagcacaca cacacaccct ttttttttta 20040
attaaaaaaa attttttttg agacagagtc tcattctgtc acccaggctg gaatgcagtg 20100
gtgcgatctt ggcccactgc aaactctgcc tcctgagctc aagggattct agtgcctcag 20160
cctcccaagt agctgggatt acaggtgtga gccaccatgc ccggctaatt tttatatttt 20220
tagtagagat agggtttcac catgttgtgg ctggtcttga actcctggcc tcaggtgatc 20280
cactcgcctt ggcctcccca ggtgctggga ttataggtgt gagccatcgc acccggccca 20340
cacacacacc ccccctttaa tggtggagtt gccaacagtt gtccgatgtg gcagagagaa 20400
ctggggagca gagaaggagc ccatcactgg gagggactga aaaccagctc ctgggcactg 20460
cagggcagag cgtgtgccgg gtgtgaaggg ggtcagtgcc cctgtgctgg ccactgtccc 20520
cctagttctg atacagttgc tgccaggaga ggctggaccg tggcaagggg ccagtgggga 20580
accggggcgg gatgggggggc agcagcagtg ctctggggtc tgcttgtggt ggccatggtg 20640
gtggcagttg cagtggcaac tgtgggccat ggcagcactt ttccgtggcg cacaggccgc 20700
tgccgggtga ggggagagcc ccgcttgttc tcctcctgcc gcctctcctc ccatccccag 20760
ctgctgggtg ctctccggcc ggcatgaagc tgtgcaggca ctgggcactc ccttgggctc 20820
ttagcggggc cttctcattg ctgcttcgct aggaaccacc cgctctcctt tcctcctgga 20880
gggccacgtg gggcgagcac ctgggacagc cgtgctgagc tctacccggc cccttcgcag 20940
gcctcgtttc tgggggcctt cagttttgga aacctccctg agggcccctc tagagtgcgg 21000
gtccgccccg cccctggtgt ggaggggaag tgccctgtcc ctgtgtgctg acaggtccct 21060
gtgtgcagca tggtcggggc acttctactc tgcaggcgcg gccgggcggg gagcggggtg 21120
gggggcgggc gggggcgtgg ggggagcggg gctggctctc ctgggtgtgg ggtggggagg 21180
cctcctcatc gccagcatgg agttaaaaac caggatgggg gagggagact tcccatttct 21240
cccgagatgt acttcatgac gcgtctggag aaaaggtgtg accagcccct tccttcccgt 21300
caaggagcgc gcttgcctgt gtaactcacc tgtaaagccc acccagggaa ggccagggcc 21360
tcatgaaaca gaaaacaaag ccgggcactg ggcccttgag ctcactagaa tttggggtat 21420
ctttccttgc accctcaccg catatggaac tcctggcgtc gtgtcccagg tcacgcctgt 21480
ccccgtgggg aggctgggcc cctcccccca gccactcccg agacttgaga cctctgcctc 21540
aggacgatgg gtgggaaggg gcttgcgggt aggagagcga gcgctgcttc cagcgggagc 21600
ctcgggggag ggtggcgggg ccgccgtggg aggagccgcg ccgcatctca ggcgcagtct 21660
ctaggggctg tgcgcatccg tgggggggac atgcgcatct caggggggct gctcgcatct 21720
gggggtgctg tgtgcatctc gggggggctg tgcccatcta gcggggtggc tgtgcgcatc 21780
tggagggggc tgtgcgcaac ccgggggggg tgttgcgcgc atctagcagg ggcggctgtg 21840
cgcatttcgg gggggggctgt gcatatctgg ggggaccgtg cttatctccg ggggcggctg 21900
tgcgcatctt gaggggtgtg tacatctcgg ggggcctgtg cgcatcttgg ggggctgtgt 21960
gcatccgcgg gggctgtgcg catctcgggg tgctgtgcgc tgctcctctg agctctgctc 22020
tttcttgcag cgtttgcctc agcatggagg gcgggggccgc ggcagccacc cccacagcac 22080
tgccttacta cgtggccttc tcccagctgc tgggcctgac cttggtggcc atgaccggcg 22140
cgtggctcgg gctgtaccga ggcggcattg cctgggagag cgacctgcag ttcaacgcgc 22200
acccccctctg catggtcata ggcctgatct tcctgcaggg aaatggtgag tcccatgggc 22260
cgctcctctt ttcccgggct tgtgggggtc cctgagaggc agtttgcagg ggtcttgtca 22320
cccctgcggt ctcttctggt tggacaaatc taagattcta gaaaagacag agacagaagc 22380
tggttcagag cctggggaga tggaatccaa acccaggctc tgaggttggt gagtggcagc 22440
tcctctccag cctggtccag cattttcacc tcgtttccac acacagcatc caaggcgggc 22500
acttcttgca gcagagggaa aggaatgagg ctccgagcca ggccctgctg caggggtggt 22560
ttgaattgag ggaaaaaaag taatcatatg tgagagtttg tatgcgtgcg tgcgtgcgcg 22620
cgcatgtgtg ttgttggtct tggacttggc agagctagag cgcctccccc tggggcagga 22680
gcaaggcagt gcttggcctc cacctgcctc caggccaggg atccaggaag cgggatctgc 22740

atccggttga cccgctcttc ctagaggtgg tcctggtgac agccattcct gagcagcaga 22800
aagctaagag gcattcctca cgtgacctgg ccttgcccac ctcttctgga cgcgggagat 22860
caggctggga tcacaaggtt ctgcttggga gccaggcgcc tgcttgctag gagtaggacc 22920
atctgcccag gcttaggtgg gggcctgtgt gaggagccac gtggctaaca ggtgaactca 22980
gaggctgctt gttgcctcag tgtgaccaac agtggacctc aaacacagcc tggaatttgc 23040
aggacacctg acaagaaggt gggtggaggg gggcgtctgc tgggccgggc agctccttta 23100
ggtggggacg agggcagagg cgctgccttc actgcctgtc ctggtggtgg ggactgtggt 23160
gcagcctccg gccctgccct ctttgtgcac agctgggggа ggttggaggg tgggggaggg 23220
gtgttaaggt gcacacagcc ctaggtgccc tcagagaagg ccaggagctg cccagggtcc 23280
ccagggaacc tggctttctc ctccggttcc tgcagctagg ctcctgttac aagccgtcag 23340
ctcacatcac ctctccacca aggtgcctcc ttgctgctgg ggcaggaggc ctcacgcaca 23400
ggcggatgca ctcacgttcc ctctctcacc tcccacagcc ctgctggttt accgtgtctt 23460
caggaacgaa gctaaacgca ccaccaaggt cctgcacggg ctgctgcaca tctttgcgct 23520
cgtcatcgcc ctggttggtg agttcccggg cgcggccttc cccgccacct gcctgcctct 23580
tcaggtatgg caaacagccg cttcacctgc tctgttccct ccccagagct gtgatgggcc 23640
cgcccccacc ccaaacatcc cctgcagggc cactaggtgc tggcagccct cgagctggga 23700
ctaaagccca gaagtcccgc tagctggttc ccctgggtca tcttacacct tccttttcca 23760
catccaaccc tgtcctcata tgacctgtct ctggcccagg cttggtggcg gtgttcgact 23820
accacaggaa gaagggctac gctgacctgt acagcctaca cagctggtgc gggatccttg 23880
tctttgtcct gtactttgtg caggtgagtc cttccaacac cccgggcctg gggccaccta 23940
ccagggaggt gggaggagga ggggagctgg gctgaaatgt acctcatgga agggcctgat 24000
ttctggggtg tgtgcaggtg agagagctgc cttcccaggc ctgtgcgggt gcaaagctag 24060
gattggcggc caagggtgcc cagctcccag gcgccccatg gctgactcag cacttttggc 24120
cacaagccca gctttctgag tgtgccaagg cctcaagcgt tgggggaccg gtcccttgtc 24180
cctaagcgct tagtgctcct cccaccactg tgctgaggag gaaggaaacc aggcagggct 24240
gaggggctca cagtcgccac gtgggcacgt gggtgtgggc ctggccgacc tcggctgctg 24300
gggtgtccgg ccgccctgtc ccggaaccgt ctcctttcaa atcctagtgg ctggtgggct 24360
tcagcttctt cctgttcccc ggagcttcat tctccctgcg gagcgctacc gcccacagca 24420
catcttcttt ggtgctacca tcttcctcct ttccgtgggc accgccctgc tgggcctgaa 24480
ggaggcactg ctgttcaacc tcgggtgagt gtcctgggtg ggagagggca gggcctgggc 24540
caccccttgca cagacctcac cctgccttca gcttccccag ctgtggcttc ctgagccgcc 24600
tctcgtggcg tcacaactgg tggctgtagt tatgcttgct aagatttggg tgcttggggc 24660
ttggctttgg ttagctttct tgattttacc ctttcaaaga aacttctggg ctatgggcac 24720
cctatttatt cccaccacgc agcaggatct gcaggacaac tgcttagagc tagaatattg 24780
atctaggttt ttacattgcc catctctttt tgtctgtgag ccatagctgg agattgctgg 24840
ttgggggcgg ggggatggtt gttctcttca ggcagggcag ggaggcaggg gctggtgctg 24900
gaatccccat ggcatcttta aggcccagga tacaaagggc atttggccta attgtgaccc 24960
cttgggcagc ttctcccttc ctctcacccc tgcaggggca agtatagcgc atttgagccc 25020
gagggtgtcc tggccaacgt gctgggcctg ctgctggcct gcttcggtgg ggcggtgctc 25080
tacatcttga cccgggccga ctggaagcgg ccttcccagg cggaagagca ggccctctcc 25140
atggacttca agacgctgac ggagggagat agccccggct cccagtgatg cgcccggccg 25200
gccctggggg ttcgcggggt gtcttcttgc ctgcccctgc tgaggcgtct tcaggactgc 25260
aggctccgga gagtggctct ggcagcaggc gggcgcgtgg gtgcagctgc atctgtttga 25320
gtgctgcttt ctggggtcag gtctccgcct cctctgcttc tcctttctcc gctgctatag 25380
accagttcat tgtgtgtggc tcccgtgtct ctgttgcccc cttcagtgca gaaggctttg 25440
ggtaggactt cgggtgttcg gtcctggtcg cagagcacag atctttaaag aagcgagaga 25500
ggaggcccca ccctcctggc agcagatgcc tggggcaagg ccaggggaaa ctgggggggc 25560
ctcagggaca ggcctggaaa ggccacgatg gctgctgaat tcaaacaagg agtccctcca 25620
gcctgaataa cacgtggcac aaatgggccc ggcctttggc agaggagcaa gtgatatgat 25680
gtgtaaagta tgttggtggt gaaagcaagg ttccccagga gaggggaggg actggccct 25740
gggaagctct gagatgaggc tgtggcccag ctgtagtcct gaccttcctc ttctttaacc 25800
ctttagccct aggatggctt tggtgggaga ggggatagaa gcccatgact tcagacagac 25860
tttctcttgg cagatgcagg cgggcctcct cccaggctgc tccagacatg ggggttgggg 25920
atgggggggca ccttgcagcc ccttcctgct ggggctccct ccttgtagca ccccccttgc 25980
ggctcagctc tggtttcctc tcccaggctc acccaggctc tgctcaggct gggaggcaga 26040
gggcacaaac cttataattt tttaaatgaa aaaccgctgc tgctggctgt ggctagagcc 26100
ccctggggct gctggagctg ctgcctctgt tctggaggac gagccttctc cttatctgct 26160
gcccatcttt ccaggaagtc aggatggagt cagaacaact acagtcatcc cccgtggtgt 26220
ctgcacatca ctccagcccc ataaagagtg tcatgttagc tgagtcacca tttggcttcg 26280
gcctggaaat agtgtgatta gaacactgat cgtgtgcgag gccaggagat caagaccatc 26340
ctgactaaca aacacagtga aaccccgtct ctactaaaaa tacaaaaaaa ttagccaggc 26400

gtggtggtgg gcgcctgtag tcccagctac ttgggaggct gaggcaggag aatggtgtga 26460
acccgggaga tggcgcttgc agtgagctga gattgcactc cagcctgggc gacaggctca 26520
aaaaagaaaa aaaaaagaac accgatcatg tgcttcttgg atctggtgac tgttctctcc 26580
ctgttttcct ttctttttgg gtgtttgagg agcatggctt agcttgagac acacacagac 26640
actgtgtact tcagtgaagg gcttaatata cagtttccaa acctgacgac tcttcttctt 26700
gtaatggctg ccctttttcct acctgaggcc gtcttagaga aaggggccag tctcctctaa 26760
tgctcagatt tcccatagtt ggcttttgct gtgtctcctg cctcaggcag tgtcatttct 26820
gggagcaggt ggttgtagtc caggcccctc cccagcaggg tctgcccagg ctccttcgag 26880
ccccttcccc cgcctcctct cagcctgtcc ggatgacagt gttcgcctcc tgtttagact 26940
gtacactctt caggggtagg ggtgccgtca gttcttcaat cagctggcac acacttgtat 27000
agtgaaatgt ttacatgtgg gaaaactccg ccttagacaa actaccaaag tacaatcgtg 27060
tctctctcta gccggaatgc tacagagaga aatggaacct tagatttgca acaaaagtct 27120
gtaaactggt ctgtttgcca aagtgaacac tggatgacta aggagctgaa gaaggccccc 27180
agaagcggat ttgtggtggg ttattttatt ttgcctgtgg ccaatcttct gtgaaataca 27240
atgtgctgtt ggtgcaacag atgattcaat aaatgtctac agcagacctc tcgcctgtta 27300
tcttccttta ctgtggtaat aaaaggagcc ggagctttta gcagccagac acatgacgtt 27360
agctctaggc ctgagagaaa ttgaatatcc gcaggctggg cacagaaggt aaagcgatta 27420
gtgatattga tcatggccta ggggctgaaa aggcccagcg gttgtccagt ctcgacccag 27480
cagaggcagt gttgtttcca catggttaga taagcccttt cctctcagcc tgagagggtg 27540
gcctggatgg tggagctgca agagcctgat aagagccttg gcaaggaagg tcccccagtg 27600
tttaatggac ccctttccct ttgaaatcag tctttttgat ccttaagaag aggagcaaag 27660
cttttggaac gagctgagat tccactttag atccacgtac gtggctcagg cagaaggtga 27720
gtttttggca aatttgggtg ggatcgtact gtgcgctttg tgccttgtca cttaaagcag 27780
tggccgccaa ccttttttggc agcaggaaca ggttttgtgg aagacaattt ttccacagac 27840
cagggctggg ggtgggggat agttttggga tgattcaagc acgttacatt tattatgcat 27900
ttttatatta ttacattgta acatataatg aaataattat acaacttacc ataatgtaga 27960
atcaatggga gccctgagct tgttttcctg aaactagaca gtctcatctg ggggtgatgg 28020
gagacagtga cagatcattc gattctccta aggagcatgc aacctagatc cctcgcatgc 28080
gtagttcata attaataggg tctgcactcc tatgagaatc taatgctgca gctgatctga 28140
caggaggcgg agctcaggtg gtcatgctca cccgctgctg ctcacctcct actgtgcagc 28200
caggttctaa cagccacaga cccatactgg tctgtggccc tggggctggg gacctctgac 28260
ttacttaaag caatttaaaa actcaccaga gctcactatt taaagggacc caacagcaaa 28320
tcctctaata aggtaacagg agtccctgct tccccgcgtt ggtctcaact ttgtcttctg 28380
cattgtagct ggtctctaac gtgtaggagc cagaacggag aggtctttcc ttagggcacc 28440
tgtagagtcc ctcctgaatc aaggcttcca tgtggacctt tatttttatt ttatttttt 28500
gagacggagt ttcgctcttg ttgcccaggc tggagtgcaa tggcgcaatc tcagctcact 28560
gcaacctccg cctcctgggt tcaagcgatt ctcctgcctc agcctcctga gtagctggga 28620
ttacaggcat gcgccaccac actcagctaa ttttgtattt ttagtagaga tgatggggtt 28680
tctccatgtt ggtcaggctg gtcttgaact ccctacctca ggtgatccac ccgccttggc 28740
ctcccaaagt gctgggatta caggtgtgag ccaccacgcc cagctgagtg tggaccttta 28800
tttactgaaa agcattcaaa gcacaaggct tgatacatgt cacaaactgg ttcctcagtc 28860
agcatctcat ttctgttttt ttttttttt ttttttttt tgcaccagcc tctgaatctt 28920
atgggtctct gtgaatattg aagtgatccc agtcaccaag ggatgatgcc tccctctcca 28980
gccagccact tcacctcttc ctctacgaat ctctggctgg ataatagcag ggacctcatt 29040
ttcttactgg ggtcccaaag cgactcctct tgactgtggc atagagttat ttctccaaac 29100
aagcacgcgg catttcccaa gccatcctcc tggacccagc gcctgcactc cgtggggcgc 29160
cattcgggtt gcttgctctt ttcagaccat cccttctgag agaggaggca gcgtggtgga 29220
gggaagagct gtcagtgtgg ctgattcaga ggcctacctg tgcctctgtg accctgagca 29280
acttcagctc tcaggatctg cttctgcatt tgttttggag acagcagtgt cgtctgtcga 29340
ggtgtgagga atacattagg gaacacatgt gaagtcccta tcacagtgtc tgggacacaa 29400
tggacctccc aagaggatgg cctgttagaa tggactctga gcaccttctc ttctatgtgg 29460
gccctctctt catcacaggg ccttccagac atgatacctg ttatcagtgc cccatcttat 29520
tcctgagaat ggaataactc agctgccaag actccaagct cctccaactt gtttttaaag 29580
ctgaaagagg gtgactccat tccccgctgg ttcgcctacc cattcccagc cgaccgcagg 29640
ggagtttggc caccatgtgc acgtgcgtac tgtgtctgcc tgcagttacc aaaaggtacc 29700
catgagcaga tatccagggg aagaaagtct gacttactta gtgcttcgga ggatctacac 29760
agacttctgc tgagtttgct gtttgcttct actctgagta gctttttatt aaggttgagg 29820
gtaaattctt tttaagaagt attaatattt taaatgaatc actccttttc ctttcctcca 29880
atgtcccagg acttcttaaa aagctggggg tttggccagg tgtggtggct catgcctgta 29940
atcccagcac tttaagaggc caaggcggga ggacagcttg aggccaggag ttcaagacca 30000
gcctgggcaa cataacaaga ctccatctct gcaaaacaaa acaaaagtta gcagggcgtg 30060

gtggcatgca ctgtggtccc aggtacttgg gaggctgagg tgggtgggtc acttgagctc 30120
aggaggttga ggctgcagtg agctatgatc acaccactgc actccagcct gggtgacaga 30180
gtgagactcc atctaaaaac aaaacaaaac atctaggtcg ggcacagtgc ttatgcctgt 30240
aattccagca ctttgggagg ctgnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30360
nnnataaaaa taaaataaaa taaaaataat tattatgcac aaagtagtag agtaagagca 30420
gtaatgagtc caggaagtgt tgttaccatc tataaccctt cttcatctga cacaaagtag 30480
agatcgctga aacctagcag aatgaaaacc atccctaaag acaatgaatt agaaagttct 30540
ggccaggcac ggtgcctcat gcctgtaacc ccagcacttt ggaaggccaa ggtgagtgga 30600
tcacttgagg tcaggagttc gagaccaacc tggccaacgt ggtgaaacct catctctact 30660
aaaaatacaa aaattagctg ggcatggtgg cagacacctg taatcccagc tacttaggag 30720
gctgaggcag gagaatccct tgaacccggg agtgggaggt tgcattgagc caagatgagc 30780
caagatcatg ccactgcact ccagcatgga tgacagagtg agactctgtc tcaaaaaaaa 30840
ggtcttttct ggccagacat ggtggcccac acctgtaatc ccggcacttt gggaggccga 30900
gtcaggcaga tcacttgagc tcaggagttt gagaccagcc tgggcgacat ggtgaaaacc 30960
tgaatctaca aaaaatacga aaaaaattac ctgggtgtgg tggtgcacac cagtggtccc 31020
agctacttgg gaggctgagg caggaggatc gcttgaaccc tggaggttga ggctacagtg 31080
agctaaaatc atgccactgc actccagcct gggcgacaaa gtgagacttt gtctcaaaaa 31140
aaaaaaaagt acttttccaa aatacttatc caggccaggg gcagtggctc acacttgtaa 31200
tcccagcact ttcggaggcc aaggtgtgtg gatcacctga ggtcaggagt tcgagaccag 31260
cctggccaac atggcgaaac ctggtctcta ctaaaaatac aaaaattagc cgggtgtggt 31320
ggtgtgcacc tgtaatccca gctactcagg aggctgaggc aggagaatca cttgaaccca 31380
ggaggtggat gttgcagtga gccaagatcg tgccactgca ctccagccag caacagagtg 31440
agactccatc tcaaaaaaca aaccaaaaaa acccgaaaat acttctccgt atcccactga 31500
aacactagtt aaggtattct aaggtaactg agagctaaat ggtgatttgc cactcctttg 31560
gcttggtggg gctggaaatc ccttttattc tcccatgttg taatgagatg tctttgctca 31620
ctctttttt tctttttctt taaatctcag aaggaggccc caaccctgag cacaacagca 31680
acctggccaa tatcttagag gtgtgtcgca gcaaacatat gcccaagtca acgattgaga 31740
cagcactgaa aatggaggtg tgtactgttt gacatgcttt ttattgatca cagcctctac 31800
cgggctcatg tctggatggc caacaaacac actgtaaatt atcagagggg ccaagtgcag 31860
tggctcacga ctataatccc agcactttgg gaggctaagg tgggtggatc acttgaggcc 31920
aggagtttgg agaccagcct ggccagcatg gtgaaactcc atctctacta aaaatacaaa 31980
aattagctgg gcgtggtggc atatgcttgt aatcccagct actggggagg ctgaggtgga 32040
ggatcgcttg aacccaggag gtagagattg cggtgagccg agattgcacc actgcactcc 32100
agcctgggcg acagagcgag actctgtctc aaaaaaaaaa aaaaaaatta tcagagagaa 32160
cggtagcaat gtagttgacg ccctcagcct agagtcttct gagagctcac cgtccttggt 32220
cattggttct aggagcacaa ctgggagtag gatgcatagg cagtatgcag ggccatacag 32280
tcaagaccgg gactgctcag ccaagttgta gggagatgag gtattataga aagaaggggc 32340
agtgtgggag aaagccctta cgttaggtgg aggaagtgtt tgtttttcag atgtttatgc 32400
cctaaaggag aaaagccagt agaaagggag aggatggggg tgatagagat ggtgggaggg 32460
gcaggtatat agagtccacg taagggtaca cctttctctg gggcaggtag ggaggctctg 32520
aggaagtaat agttatctta atagctcaaa attattgagt gcttcctgca ggctaagcat 32580
ggataaaaac tctatgtgtg ttatctcatc gaatcctcaa aaccaccgta agtagaaact 32640
gaggcacagg aatgttaggt gagttgactc aagtcgcata gaaaatgaca gacttgaaac 32700
cagatagtca gactttagag cctgtgctct aaccgttatg cttattgcct tcatgtaggc 32760
catggtgcag agaaggtggt ggagacagca actattttcc ccacctagtg gccctttta 32820
cctttgtgaa agaggaaggg agaagttatg agttgagaat tcctggaaag ttgcaggaca 32880
gggtgcttgg agtggggaat attgaataca tggacaaggg atgagtaggg gcatggaagg 32940
tccagaagag attagtccct agagctctgg attaattgtt tcaattttga gaccatgctg 33000
ttttcagcag ctcttacgac aaggcctagg tgtgtgatat ttcacagcag tactcagcag 33060
cccagggtag ggaccttgag acttgaccgg tgggtttctt acaggatggg ggtgtaggag 33120
gatgagtctg agtgggattg ggtagggact agatgtggca tttggttagg gatggtagga 33180
agcaaagctg agagatggct gatggattgt tcagggaccg gagaaacaga ctccttccaa 33240
gttaggggta gtacttcctg ttacctcaaa agttggttag agggtgggga caaaaataaa 33300
gattgagagc tcaggctcca gcctagagag cccatcccat ggagttagtt ttgcgatact 33360
tgtagttctg ggctgacatg tgggaaggaa ttgggaaaca ggaaaatgtg cttgtgttgt 33420
ttattgcaga aatccaagga cacttatttg ctgtatgagg gtcgaggccc tggtggctct 33480
tctctgctca tcgaggcatt atctaacagt agccacaagt gccaagcaga cattagacat 33540
atcctgaata agaatgggta agtgtgcgtc tgggaggagt ggtaggggac agagccttta 33600
tgttccaatt ctctgcaagg caagtactgt tgatctctgc tagtgtttca gggttttgtt 33660
ttttgttttt ttgtttttttg agacggagtc tcactcttgt cgcccaggct ggagtgcaat 33720

ggcatgatct tggctcactg caacctctgc tttctgggtt caagtgattc tcctgcctca 33780
gcctcctgag tagctgggat taaagatacc cacagccatg cccagctaat ttttgtattt 33840
ttagtagagg tgacagggtt tcatcatgtt ggccagactg gtctcgaact cctgatctca 33900
ggtgatccac ctgcctcagc ctcccaaagt gctgggagcc accatgcctg gccatgtttc 33960
aggttttaag cacacttgct ccttagcaga atctaatgag ttaattgact ttcccttaat 34020
gtagtttcta ctagcaggat cccaaagact tgttcatcgg atgtagaagg ggatctttgt 34080
cctattatcc ccatctgtag ccatataatg cctaagtctt aaaatgctcc accaggactg 34140
gcatcttaat ggcacaggaa attatcagct aagtctgctt cctcccaact gcaatcaggg 34200
cagaaaatcg agcgtgggga cattcttggg ctgttccctt acccagcagc agctgcattt 34260
tctccttgag gatgcgatct gctccatgtc tgtgtgtact tcatagactg ggtaagaaag 34320
agacagtgat agaatgatga gctttatgga ggctgagctt ctagaggcat ccactcatgc 34380
cagcctgttt ccttccctgt cagaggagtg atggctgtag gagctcgtca ctcttttgac 34440
aaaaaggggg tgattgtggt tgaagtggag gacagagaga agaaggctgt gaacctagag 34500
cgtgccctgg agatggcaat cgaagcagga gctgaggatg tcaaggaaac tgaagatgaa 34560
gaagaaagga acgtttttaa agtaagcatg aaaacatggg tgtttggtgg gcttccggga 34620
ggaccctgat agatgcctta tgcatgcctc ttggtttctt ccttcatgtg ccccagggta 34680
taggtgagac agttcacata ttgtacaaac atttattaag tgaatactga ataggaccng 34740
actctagtga ggactatgag aaagaaagag tcaaaancat tgaaaacgat gtccctgcca 34800
gtgaagagct caaacccagt gatccattcc agtactgaca ttcaggactt tgcaaggtcc 34860
tgtgtggctt tgtctcatta cctcctggct cctcacctcc tgactttcct ttatccctag 34920
tttatttgtg atgcctcttc actgcaccaa gtgaggaaga agctggactc cctgggcctg 34980
tgttctgtgt cctgtgcact agagttcatc cccaactcaa aggtgcagct ggctgagccc 35040
gacctggaac aggccgcaca tctcattcag gctctcagca accacgagga tgtgattcac 35100
gtctatgata acattgaata accaggctac atgtgccccc gggttccttc ctagaaatgt 35160
ggcagcccat tccagcacac aggcttctgc agcaatctct gagggtaaag ccggtgggag 35220
gctcagcnag ccaggaggcc caaggacagg acttgcgacc ttgaagccaa aggaatctca 35280
cttgtggggc ctccttgtca gctctgctgc tgtctcagag ccatctggat gagtgtcccg 35340
acaccctctc ggatgcaggg caggaccacc cagctggtca gactctgatg ttgggtagct 35400
ggcctctgtg gggattgtaa gtgccctgag gcgctctgta ctagaaactg ctcttaataa 35460
taacggtgat tattggttgc tgcattgctg ttgtatggct cttgagtctt cctgagtttg 35520
tgtccagctg ttgggatcct ctggactaac tttcaagtcc ctaggcttag ctctactact 35580
ccaatcccag gattattttt ttaattggcc atagaaatgt agttgtatcc aggcactcca 35640
tggtctgatg ctggtcttac agcctaaacc tgtcattcat tcatgcaact aatatttctg 35700
gaacactagc tatataccaa gaactgtact tatctagagc ataagctatt caactggtgt 35760
cctgtgttag atcctgctct agaaggtagg acctgaacac aggactgctc tttactggag 35820
ctgtcccagg gcagtgtgag ctgctgtgca agcatctgtt gctctctttg tgtgacttct 35880
gtgcccttct ttcaccaaac agcttaaggt tctagcccag taggaaatca agtgccagtg 35940
ataagggact tggcagtggc tcacacctat aatcccatct actcgagagg ctgaggcggg 36000
aggatcactt gagcccagga gtctaaggct acagtgagct atgattgcac cattgcattc 36060
cagcttgggt gacagagtga gaactcttct ttttttttct ttttttttct ttttgaggn 36120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 36180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnng ctgtgtgttg cctgcagttt 36240
gtctgcttcc ctaggctttg ggaaatgcag gtatgcacag caggaatgtt tgtttagcac 36300
aaatctgatg tcaagaggaa gtggtctccc tcctcgcttg tccctgaagt aaacacagag 36360
gtagcagtta aacaggaaag gctgcaggct cagtgctgct cccccgcttc cccactccct 36420
tttagatggt attgagtctc cagccagttg caggaaacct gccacttttc cattcacaat 36480
gcagctgcct agctctctat atggtgagga ctggagttcg accagatttc tgatgaaaag 36540
acaaagcccc tggaagcagt gggcctccag ttgtgacgaa gcccttaggt ctgcagtggg 36600
agctgtgccc tctgctcgct accctctcag catgcagcct ctaaacacag ccactcaagc 36660
tggcagccca cctcagcctt ggctttggta gggtggctgg attgggtgca gcttcagcta 36720
caaaatgaaa caagagctca gtgctccctg ggtatagcta gagagaaacc tcttccaatg 36780
aagaaaaagc ccccctttct caacaaatgt cacaggcctg tgctgacagc tccttagaac 36840
ggcgtctctt gtccacagct ctcttgaggg cctaccacag tgtccaactc tgaggagagg 36900
aggtgctcaa taacgtgacc gagcaaataa atgcacaaag gggtggctct gaatgtccca 36960
caagagacca agcaaatcat cgggaatttg gccttagccc taacctgccc aattcctcct 37020
gagggagggt gcccccatca ctgcatcctc aggagggtaa gcttaaaagc cagctccctg 37080
aagaggctgt aatggaagga gaaaagcatc acaacaccat ggttttccaa gtgttagcca 37140
tttataaata agtacatttg ctttcataca tacagttcct tgtacagatg acaatctgta 37200
tacatggggc aggaaaatgc attcatttga acttttcaca tctatctcac acagctcaca 37260
tgtacagaca ataaaactgc tcaagcaagt acagcaaagg aaaatgtctt tccttataca 37320
caggggtaga tgcctctgtg gggtgtgggg catcccactg cacggcttca caactgtgtg 37380

gtgttcaata tatcaggaga gagaacaaac atgcattgga taatatactg tacagagaaa 37440
gtcctttaca tctgagtcat agaaaaccta aaggaaaact aagtgcatta aagcttttc 37500
cagcaagtgt cttgaaagga cagcaaagag gaggaagaat caaaatcata ttagtacaaa 37560
tcactcttta attgtagact gtacatgtct gtactaatta aaatcatctt ggatttggag 37620
gagacagaac agagacaaag atgctgtgct agatggaaag gaggccacgc ctgaaaaggc 37680
acctgccctg agcctgatga ggaactggcc tcactcagca ggaatcagcc aaaggaaaca 37740
aaaaacaaaa caaaaccaga acaggaagtg taacttacag gatttccaaa atcacctgtg 37800
aatgaagtgg agatctggag ccggcatctc ttaacttttt ttttcccccc agtaaattgg 37860
tatgcaataa ggcaggtaca ttcaagtact gaattttcca gaattaactc ttgtctggcg 37920
ctggggacca aagggattga gttgagcccc ctctaaccag actttctggt tagcgattag 37980
caaaagaaaa attcagccag caagtgctac aaaaacaaag cagctagggc acttccgttc 38040
cacagagtag gtctacctgg aaaaatgagc gcggcgctgg cctgatctct acgcgtccag 38100
cggcagcctg gcaagtcagc tcagcgtcgg tatcagagtc agcaggaggc aatgagatga 38160
tggggtgagg aaacatgaaa gtaacacttg atttttggtg tccaattatg cgttcatttg 38220
gtactgactt tcaaagctct gactgtggcc accatgtggc cacaagcatc tcagggtggc 38280
tcacctgctg tgaggctttg gacaccgaaa tgaaggttac caacacttca gcccttgagt 38340
ggtctgtatg acaaaaggag ttgatgaaaa cccagtgatt attcaagtag ctctgcacag 38400
tggctccacc agccccattg tgcttgtgtc cagcccccag ccagccacct tttcttggga 38460
gcagccagag ctgagttcaa ggcattgcat ggtgaagggt tccatgacac gtctttgcag 38520
gtagctcttg cccctaagcc ctttgttcat tgttgttagt catggtagat ggtcggtctg 38580
gaattcctag aggaagagga gaaagagctg cacctcccag tgagcgagca gcaggaccgc 38640
acagccctcg gtgtggaagc agcagccggc tgcctttgca ggtcggtttc ttgagaaatg 38700
ggcagcccag gaacagacag gcagggtgtg cgaggctgct gggcactggg agcatcaaga 38760
ggaggctggg cacaggggcg ggcacctctt gccccctggg atagcctatt ccattttgtg 38820
gcaaagattc agtgagcact ggttttgtcc aaggcatttc tgcagtagaa aaataactct 38880
ctctgaatca aaccacccaa ctgtgacgct tctggagtta taaaagctgg aggctgagag 38940
gaactgacag gagggaggct catggggaag aaggggcttc tccatgtacc catccctcta 39000
ccagagcaag ggagctctgg tcaaccttct tccagcctct gcctggctga agtccaaagc 39060
atgtagtgtt caaagagttc gtcttgcaca actggcacag atgcacgaag acccccttcca 39120
ggcctctgtt gccttctcct gaacctctga gcccgcttgc ctgttgccct ggaagcagct 39180
cctctctccc gaggggatgc cgctgttctt tgctacaaaa caagcgcagt gcagagccac 39240
agtgacacct agtggtaaac tatggtgaag cacaagtgac atccacatag cccactgtac 39300
gtgactaaaa tctaaggaaa aatacttatg gatattaaat tagatactga ttaatttta 39360
atttttcta ttgggtacat ctctggaata taaaaatacc aatatttaga gaggggctca 39420
taaactacta tacaatatta aggactgaga ataccttct ctaagctgtt ctgtttacaa 39480
taatttagga aaagtgttta ataatccagg cttaactaca ttagcaaact tgtatatctg 39540
gacaacgacc tggggtactg tacatgattc taattagtgg aattttcctg agccatcgag 39600
tttaagttat acacatctaa aaagaggggg cacatggggg aggagcgaga agggtttgtg 39660
ctctattaac ttgggacttt aatagtgcac gtctgcaacc cggacaagat acaacagcag 39720
atacaaaatg gtctccattt tgtcatgcca aattctcatg ttacacaggt tttcccttta 39780
ctttgtaaat aaacattaat tgttaattgt taattgtgtg ccttactgat ccagtagata 39840
aagtaaccgt gtctcaggag tccctaagaa cattgctgga aaagcacttt aaaatcactg 39900
caaatatttt tcatattaaa aaattcttaa tctttttgat gcttatatac aagttatttc 39960
ttgtgctata aatgttgtga tccactgctt gatgtctttc ctttcctttt ttcttgaaaa 40020
atacactaaa agacaagagc ggttctgcta ttttctaatg aagacattac tcacacttaa 40080
atatccagta cttcagttac aaattcaaac agtaaagtgc acccatttat agacatgatg 40140
tgatagaaac ccattagtgc aagaatcctg ggccaatgga acatacaact tggtgagaaa 40200
cctattaaac tgaagtttgt cacctctgtc ctccnnnnnn nnnnnnnnnn nnnnnnnnn 40260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnn 40320
nnnnnnnnnn nnnncagggt gaaagataat gtgttctgga gtcagttctg gtttcaaagc 40380
atgactttgc ttcttgctag atgaatgatg cacaacgaca ccacctctct tgccttcagt 40440
ttcataattt gtaaaaggca gataataata tccactttgc aaagttaaga gctattatta 40500
taatatgaaa cactgtaaga gtactcatca gagagcacct cctgtagtag taaatgcttc 40560
acattttcag ttcagtcccc tactccctca gggtttgctt tgggactaaa aaccactgcc 40620
acataggtaa gcatggaaag aataagctaa acaccaagtg cagagaaaaa gactcaaggg 40680
tcaaagtgga ccacaggctg accagaagca tgccaagagg ctactgtgaa agtaacctgg 40740
atcagcctgg gaactggatt tgggcacctc caagaagagg tgggnagana acagactgca 40800
ctgaagggtg tgccaagtgc tgctgagagg gtaggtcagc accttgtaag gaaaggataa 40860
gggggattcg ganncnggaa ggctgaagcc tgaggtagaa ccccagtgct ggcttcctat 40920
ctngggacag tttggcttac tggcatatgt gtgtgcatgt gtgtgtgtgt gtgtgttagt 40980
gggaggggct aaacattgtc ttttctttct ttctttttt ttttttgag atggagtttc 41040

accttgttgc ccaggctgga gtgcaatggt gcagtctcgg ctcactgcaa catccacctc 41100
ccacattcaa gtgattctcc tgcctcagcc tctcaagtag ctgggattac aagtgcctgc 41160
catcacgccc ggctaatttt tgtatttttg gtagagacgg ggtttcacta tgttggccag 41220
gctggtttcg aactcctgcc ctcaggtgat ccgcctgcct cgacttccca aagtgctggg 41280
attacatgca tgagccactg cgcccagcct aagcagtgtc tctgaacaca gaaggctgca 41340
caagaggaaa tgggctctga ttaatgtagg aaagactgca gtctgacatg aggaagaaat 41400
tccgctagaa tataagctta atgaagctct gctcaaggtg tctagtgtag tcaggtgtgg 41460
agagggagcc tggagcagga cttccttaaa gggagatggg ggagcccaga acaaatgaag 41520
gagagggcac tcaccccagg agcttcccgg caatcttgga gttgtagatg tcacactggt 41580
gcagtggacc catgtggccc gaggctttgc acagtgtttc gtggaactgg aactggagca 41640
ccagactgag gaagtacctg gttgagaaac gtgagggcag gctggggagt ggagaccagg 41700
agtggcaaac tcagagccac ttgggaccag gcagctctga catgcagtgg ccacgagggc 41760
cccatgtcag cagtcactgc ctgagcggag gctccttggg cagggccagt gcttgggctg 41820
accctagagt gctgggggggc tctccttctc tagctttctg aaccagcagt caccccaacc 41880
cttgcccccc tctgacctgc tctcctccaa gggacagcag ggatggtgta gaatattctc 41940
atcagctctg ccctctacac tggacctgat gcaaaagctc aaaggattct aggtcagagg 42000
cctcagaagg ctgcgcagcc caacaggaag agatgggcac gggtacaggg tgacatttgg 42060
acctagggcc ttcctgcaaa tcctcacccc agttcccatg cccctcaact cctaggactg 42120
gcactgctaa cagggcccag acctttaccg tatgtagggc acaccagcag aaaagtggaa 42180
cttggcacct ggatcaaagt cttcctctga gtgaggaata gcggggcaca agccctggta 42240
tttcaacctg gcagggaaaa ggcagagggc ttggtggtgg tgtctggcag gttctcagag 42300
cctctttctt ttccttcctt ccttccttcc ttccttcctt ccttccttcc ttccttccct 42360
ccctccctcc ctccctccct ccctccctcc tttcttcctt tccttctttt cttctttctt 42420
ttttttttt tttcagagtc ttgccctgtt gcccaggctg aagttcagtg gcgtgatctc 42480
agctcactgt aacctcccca tcctgggttc aagtgcgtct cttgtctcag cttcccaagt 42540
agctgggatt ataggcgtga gccacncgca cctggcctgc ctttttcttt tcaaagcacc 42600
tacgtgtcta tcaaaactgg acttggcagg ctgaggatac ctaatgcatt tttattgatt 42660
aaataaatga aaacacatgg gattgggaat cagagtaaaa caatagcagt taacactgac 42720
tgaacacttc ctctatgctc agaaccatgt ctggctctcc acatagactc acgcatcatg 42780
cactcaacac cgtcctggtc tctgggggag ggaccagtgg gggagcgggg gaggcaaggc 42840
ccctgcttgc atcctaacag tgggggacag attatagacc aataaacaag gtcatttcat 42900
ctagtaagaa gtcttgagaa gaaaaaagag taacatcact gtggggcagt actactatta 42960
ttcccatttt gcagatgagg aaaccaaggt ttggagatca gggaattggc ctcaggacac 43020
agagccagtc acagtcacac cctgataacc tgaaggtctg tttttgcccc tctgagatgt 43080
gcgactgtgg gaaagttgct gattgtctcc gagtcttgtt gtaaaatgtc tataaaatgg 43140
aaggaagtag gccaggtgtg gtggctcaca cctgtaatcc cagcactttg agaggctgat 43200
gtgggaggat tgcttgagcc caggagttcg agaccggcct gggcaacaaa gtgagaccct 43260
gtccccacaa aaaatacaaa aatgagccag gtgttctggt gcgcatctgt agtcccagct 43320
actcaggtgc tgaggcagga ggatcgcttg agcccaggag gtcagggctg cagtgagcca 43380
tgattgcacc actgcactcc agcctgggca acagagtgag accctgtctc aaaaacaagg 43440
cttcatcatt ggctcactgg agcctctgaa ccatctccaa ggtcaacaga gtgtcaggag 43500
cttccctccc agttgacagg gccttattca aggcccaggg agttgccacc cctcctgctt 43560
agccccaggg ttccatcagc tggggtgatg ggcagggctg acagaggcct cctgcaggtg 43620
cctcagatgg cctatgaggg tgaccttaga gaggcacgag gtcaagactc tgccaagccc 43680
accagggaga cctgctcctg aagtcaaggg tgggagtgag aggacaaggg gaggagctcc 43740
cagctctgtc acctgctcgt ttgctgtgtg acctcgggct ggtttgtgac ctctgtgagc 43800
ttcagctgtt gcttctgtga gctgagcgct ttgaagtgat cacaacgcaa ggtggttgta 43860
aatgttcaaa ttaaacataa gaggccgggc acagtggctc atgcctgtaa tccccgcatt 43920
ttgggaggcc taggcaggca gatcccttga gcccaggagt ttaaacctgg gcaacatggc 43980
gaaccccatc tcttaaaaaa aaattacaca aaactagcca gtatagtggc gcacacctat 44040
agtcccagct actcaggagg ctgaggaagg aggattggtt gaggctggga ggtcagggct 44100
gcagtgagcc gtgatcatgc cactatactc cagcctgagt gacagagcga ggccctgtct 44160
caaaaaacaa acaaaaaaag taagagcctg ttgcagagcc aggagcaccc cctggcaacc 44220
ccgaatcaca ctatgaccag cagactgaag ctggtcagtc cgcggcctgc cacatatgga 44280
ggagccctgg tcacaggcat ttattatccc atcaaatggc caggggtcag tgatgaggtg 44340
taacacgggg gtccctcact cgagagtgat gcctcccaac tgctggtggg ttcccatggc 44400
agcttgggga tgttaagccc gaggcaggca tgcgtgcatt ttaggggcct gcctctctcc 44460
catggggagt ctttccccag aaaggaaagt gctcaggccc ccgcacactc cattgcctgg 44520
caccaggaag ggagggaggt ttgggagggc ctgcctggtg gtcagtgttc tgtgattaga 44580
gaagctcagt tgcaatcatt gtgtgagctt cccagctacc cagggagcag gtggggccat 44640
caccagggtc catttttcac agatgaggag actgaagcct agaaaggcta aaagacctgc 44700

tcagtgcctg tgtggaagat tagagaggtt cttgtgctgc agaaagtcct ggaaatggcc 44760
catctggtgg aagatggtgc ccaacccacc tgaggttcca ccactcctga ttgtagatgt 44820
ccttccagat ggtgccgtca aagaccttcc agcgaaacag gtccatcagg tagccaaagg 44880
ggatgaaggc gatcttctcc agggcaatat gcatcaggaa attgacctca tcctctggga 44940
agagggtgag ggcagtcaag aggcaacccc tgggatctcc ttcccgtagc tttgcccgag 45000
agccctgacc ccaccaagcc ctagctgcgg tcccaggctt gtgctgccct cctgctgggc 45060
ccttgatcct catcacctga gtcctggtgc tagaggctga gcaggcctat gttgagcagg 45120
tgcttgtggg aggaggccga gagggtgatc acagacccca cagcctcttc aaaggctggg 45180
ttggcacctg tgcggaagat gatggagagg ttcttgtact gcaggaagta ctggaatggc 45240
ccatttcgtg gaagatggag agcgggtctt ctgtggtcac ttccgcgcac ttctttattc 45300
tacaagcaaa gggcatgggt ncaaggagca tggagaaagg gagtttcccc attcagactc 45360
cttgagggat tgggtggtcg ggttggagcc cagacaggcc ttggcagttg ttctttttt 45420
tgtttttctt ttgagacgga gtcttgctgt gtctcccagc tggagtgcaa tggtgcgatc 45480
ttggctcanc tgcagcctcc ggctcccagg ttccagtgat tctcctgcct cagcctccta 45540
gctgggatta caggcacatg tcaccatgcc tggctaattt ttgtattttt agtagagaca 45600
gggtttcacc atgttggcca ggctggtctc gaactcctga cctcacgtga tccactgncc 45660
ttggcctccc aaagtgctag gattacaggc gtgancaccc atgncccggc cggcaattgt 45720
tctttcttac tccaggcccc atatccaccc agcaaccaag ccctttcctc tcacgtcacc 45780
tttccttctg tgtttctcac ttgccttctc cctcctctgc cctggctttc ccctgcccat 45840
acctaagaca gccagacagg gcagaggaga ggaccgggtg tgtggtgccg ctgtgagcac 45900
ctgaaatcgt cgtcctggta gaagttccag gcagagatgt gacactccac ctctcgccca 45960
tcggttggcc tcattagcat caacttttte cagaaactgg gtggggcagg tggcagcgcc 46020
aaggnaggtg aagaatgtct cagcctcttc caacattttc tctggcttcc agtgctatgg 46080
gagaagaggg ggtgtggggg gcgctgccct gtttgtccaa ggatgccaat ggggaggggc 46140
agggacagac aggccaagca ctaacctgga ctttcatgat ctttgtgaca tcctctggga 46200
tcttcttcag gaagggcagg accgggtcta agatgttgac ccaggactga gccaacgtgt 46260
tctctggaag aggaggagag gggctaagag gaggctggag acagtctctg cccttatctg 46320
gcactcacct cggccaccag cagggccttt acccaggagg tgggcaggga tgggcccccc 46380
caggtcgatg agctcgggcc catagtggcg gtggagggcc ctgcgcacgt aggtgtgcgg 46440
gttcaggtag agtggccgca gctcctggaa tagccgctcc aggtcttgct ccagggtatc 46500
cgactcatac ttggagtgcc acaaggcccc catgtctttg tnaacctagg acaggagaga 46560
ggactcacca ggagctcacc atctcaccct ttagccatgg cctagctgac aggataccca 46620
gacgccttct agggaagcca ctcaacctcg ctgaccgtct gggtcttctt tggcaaaacg 46680
gggagaatac ctgcccattt cagaaaggca ctgagcaaag tccatgatcc gagctcccca 46740
ccatgcctgc aaggccctgn agtcccctct tatcccacct catccttagc cctgtggcca 46800
cctgggcttc tttcaggtct tccagagcac cggctcctcc tgccgcggag cctcntgcct 46860
aggctgctct tctttcacat cttgcttccc gtcttttgcc catctttcag gtcttggata 46920
aatgtcagag tnccttccct gtccacttgt ctgcagcagc ctcttctcat tctctttctc 46980
agcagcctct tcctattccg catcacccca gatggtagtg agagacaggt gtttactgaa 47040
tatggtctgt ctccctgcag gaccgtaggt gcatngangg gcaggtgcct gtcttttcac 47100
ctatgtgtcg tcagaacctc cccagcacct agcacctagg tgnctctatg aatatttgga 47160
gagcccagct gcagggctca taggtcctgt ctccnagggc cctacggcca ccttcactct 47220
gggaaagctc agctttgccc cttccttgcc atcactccaa gctctgtggc atcccatctg 47280
ctcaccgttg agctgtgcag ccttgttgct gagctccaca tagtgctcga aggtggtgca 47340
gatctggcgg cccacngcat cctgccagcc ctgccaggcc cacagcagct cctctttgtc 47400
cctggaggtg gccatgactt cgagttctat gcccagaaaa ggagccatgg gggacccact 47460
cccacttgcc tcttccctca ccctcctggc aataagggag ggtggttggg ggctgggagg 47520
ggcccctgga agggactgag tgctggagat tctgtggttg ggtgggacag ggcttgggtg 47580
ttcaccagac tccagggaca ggcagggccc ctcattcagg cacacctggg ccatactgta 47640
tgtcatctcc aggntaggcc agaagctcgt tatactgggg ggacaggtgt tacccagggt 47700
aggctggtgc cccaggtccc cccccagtgc tcgccccacc agccaggaga gatcggtacc 47760
cccacctcta ctcctctcct tctccagagg aagttctgat cctgcccttc ctccactccc 47820
tgcctccctc ctgccccagg cactctgtct cccctaccag acccctagga ccagggagca 47880
tgtgncctaa aggcctgggg gttccatcca tcacctcccg cagctcgtcc ttggacagag 47940
ccgccttgtc tatgttctgc agcttactca gcatgccatt cacatccggg tccttgaact 48000
gggtgacttt aaacaggtgg gcctgggtgc caaagtaaat catgaactgg gacctctcca 48060
tgtccttgtg cagctgcaaa gagagccacc aaggttgggg gacaggggaa gggtcccact 48120
tgacctccag ggccatggca ccctaactga ctccctaatc ccattggctt agggagcaag 48180
atttgctgag agggagaggg cagaaacctg gatctctaaa agcaaggagg tgagatttac 48240
aaaggacaag aagcaagaga ttcttgcctg gaggaagatt agggcttcca ggaggcccag 48300
agcttgcagc cctctttgac atgtacactg gggcttccct tgttcctctt gcttgggatg 48360

ggggcaggag gtgtggtgcc ccatggtccc aaatccctaa aaagagcaaa gagaggagag 48420
gctggggtgg aggtggtatc acatcatctc ctcctgattt ttcctggtga tattggtgac 48480
atagttccaa gtggcctcca tgaacttgtt caacacaact tcacctgttt ggtcataaaa 48540
ctgcaggaag atcttggtct cggtctcatt gtagaagtca tctgcaggga aggatatggg 48600
catcatgtgc tgagaaggtc ctgtccgcac ccagtccctt tgtctggtcc catgcttctt 48660
ggggtcccag ccacacctaa actgtgttca cccttgatcc tgagccaagg caagagctgc 48720
ccataacaga ggagcacaag tagggaaggt ccagggcaag tccaccttgc tcccatgacc 48780
aacagagcag cagagctggg tgaggccagc tccatattgt gacgtcagag gccagaggct 48840
cagctacaga atgttagaac tggaaatccc catccacctc ctggtccaac tgtccccatt 48900
ttacaggtgg ggagactgag gtccagaaag tcaagccaaa tgatctggtt catgtagtta 48960
gctaatggcg agccaaaaag aggaaccagg ggtcttgact tttaggccag aacttgttct 49020
gctctcctcg gccacttttc cctggggctt gagccaccct tgggttatgg ccagggattg 49080
atgaggccct ccctccatag gctgggacca cttctgagca ggtggatggg atgatggatg 49140
gatggatgga tggatggatg gatggatgga tggatgtgat atatctgggt ctctggaagc 49200
caggttatct ctctaggttc ttctaacttg acctcttatg ttcacatgtt tgggcttttt 49260
ctctcttatg ggctgtgcaa tcaaccccag ggctgtgcag ggccctcccc atcataccta 49320
tttctccagg gagaccctga tgttatgagt ttgcaattag gagagggcac acaaggtagg 49380
gactgggtgg ggtgtggaga ggtgaggcgg gcaccctcca tnccagggtg catgcccatg 49440
agtacctgag cctgatctga aaaccagctg gcaagacacc agctctctgt aggcctgtgg 49500
tcacccccag tactcttgca ctgaggtctg tggaggccag ctccatctgt ctcaccttgg 49560
atttctcctc agactcactt ccctccacac tcagaatccc tgtaccctgg gttgccttcc 49620
atcactctca gcccacaatg gctcacatct tcaaaaaggc cctacctcta cacatgggct 49680
gataccacga ctagagaaac tgaccaaata acccccccaat atgatccatt atgggtaaaa 49740
tatatttata acatgaggaa agaaagaaaa aaatagtgac acttctggtt aagctgatcg 49800
actgagctca caggaaagtc tcccttcctg cttcaagcat gtagacagtc cattctcaca 49860
ttgctataga gaactacctg agactgggta atttataaag aaaagaggtt tagttggctc 49920
tacagttcca caggctgtac aggaagcacg gctggggaag cctcaggaaa cttacaatca 49980
tggtggaagg cgaaagggaa gcaggcacat cttcacatgg tggcaggaga gagagagaag 50040
ggggaggcgc tacacacttt taaacaacca gatcctgaga actcagtcac tatcacgaga 50100
acggcaaagg agaagtctgc ctccatgatt caacacgccc acctctcctc taacacaggg 50160
gaggtcaggg tgcatccatc cagagagctg cagcattggt gtgggggggat agctgatccc 50220
gatggaaatg gcgtctctgc caagctggga gcccctgcaa gcgtcaggtg catgcatggg 50280
attgtgcagg ccattccaac ccaactgaag atggctgata ggtacagttt accaagtata 50340
agagaaaatc cagttttagg agggacaata ataattacca accctcactg tgggttggtt 50400
ccaggtgcca gggaccatct gggggctttc aacatatgaa tcctcacaag aaacccatga 50460
ggtaagtact gtgatcaacg ctcccatttt acagatgagg aaagagaggt acagaacggt 50520
tgaggaattt gctcatagtc acacagccag cagatggtgg aagtgggact caaattcata 50580
tggttggttc cagagcttgt gctaaactgc agacacaatg aatgaagcag actgcaatgg 50640
aggaaggaca agtagcagag caatccaaaa gggacttaaa atgcaggtct ttatgccatg 50700
tataaaaatt agctcaaaaa tggatcacag cttattttct tttttcaatt taaaagagta 50760
catttaaggc tgggcacaat ggctcatgtc tgtaatccca gcactttggg aggctgaggc 50820
aggtggatca cttgaggtca ggagttcaag accagcctga ccaacatggt gaaatcctgt 50880
ctctactaaa aatacaaaaa ttagccaggc atggtggcac atgcctctaa tcccagctac 50940
tcgggaggct gaggctggag aatcgcttca cctgggaggc agagtgcagt gagccagatc 51000
acaccactac actcnagcct gggtgacaga atgaaactct gtctcaaaaa ataaataaat 51060
aaaataaaaa taataaaata gtacatttaa caccttatct tggacaacgt gacagtcatg 51120
agatctttgt gtnggagaag aanggggggaa caaaagagca aagtctacaa aattcttaga 51180
acacagaggg aagtttcatg acattggatt tggcaatgat tgcttggata tgacatcaaa 51240
agcacaggca acaaaagaaa aaataagcaa attgaacctc atcaaaattt taaacttttn 51300
catatnaaaa aacactgtaa cagagcaaaa agganacaac ntggaatggg agaaaatatc 51360
tctgactctc agatcnacga cgaatnaanc gccagaaagc aaaaagaaat nnnnnnnnnn 51420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 51480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gaaagaatta cccccgctca ttcttgtcta 51540
aagccccccc cttgcacctc ttctctcccc cagtcccgga gctggtaacg gccctggccc 51600
catgagcagt ttgccttctt gagtcactgc ctgtgtagta catacctgac cgggagtcca 51660
aaccaccttg gtgctctgaa gtccactgac tcatcacacc tttcttagcc tggctcctct 51720
caagggcatt ctgggcttgt aaacagacat aggaagcctc tgtttaccct gaagcaccac 51780
tgtccagccc attggttccc actggcagca tggtagagct gagagaaaca ggctctcagg 51840
gtacctgact tgaggggaat cgtttcatga agctgaactt caagcatatt tccagtacat 51900
tctttcagag tctgtttttc catccaaata taagccccag gccattccac ttagtgtctt 51960
ttcaatgata ggcaagaatg atatctgagt tgaacttcgg tgcttctgtt gtttgagttt 52020

actgtgcctg gtggtatatt gggcattctt tggattgagt gttctgaggt gagagagtct 52080
tcccgaggca tcctgtctgt gcttccaacc ctgaacaaga ccttacatga gagatggact 52140
gatggactgc ggcaatcctg ggctgtcaag tggatagata gttaaaaagc attatactgt 52200
gggtaatgaa aagggagagg aaaaaaaaag aaggaaaagg aattatagac ccccagggtc 52260
agnccagtta agagctctac ccacacctgt caacccctct ctcccccagt ttaggttctg 52320
agcagtattg gacttgtagc ctgcagttgt cttttgactt gcaggccgca ggtgtctttc 52380
tgttatgtga atgagttcca tggaggggca tatgtgtgat tccaccgtta gatgagccct 52440
tggggcagng cagtttggga tgtgctcttg ggggaaagtt ggctgtttcc ttgcgctctg 52500
ctcctacccg aaggttttta agtccctctg aattgnctnc atctgagatt agtagagtag 52560
caggcctgaa ggatgatggt tttgtcctct ttggttctca cctgcttgag aagtaaaaca 52620
gtaactttgt tcttctgggc ccttaagctt ttttggttaa gtcttccttt tcagaagtag 52680
atgtcattat atgccaaaag tctagctctt tgctttacca tacagggacc tgtcccaaag 52740
aaaaaggctc tttttttagc cagcatattt ccccttctac cctttтactt tgttgttctg 52800
attttaggac tctggctggc catgtgcttg tggttgcctc tcctgcattt gccactggat 52860
ttgcactgca tcgtttggag atacaaagcg agcagttctt ggtcagaacc ctcctctgct 52920
tttcattgtg tttgataatg gttactgggt ccttctctca agggtagcaa ggccaagctg 52980
atggctgctt gtttaggagg ccatcagttc cttcctgtgg agaagggtct gaaatggaag 53040
tcagtggtag aaggggctgg tctgctgggc agggcttaca tccactgagt tctaagattc 53100
ctttcctgat ctgcacctac gcctggtctg tatggtggaa tttgtcagct ggaactcaga 53160
aacaacaact tgaaaaaaaa ataataatta gaacatattt gcataagata gctatttact 53220
ctggaaacca acaacttttg agatttccct tgccctgtgg acgcccagct cctgtcatcc 53280
ttccttaggt cctgcagtac agtcttcccc tgaatgccac cggggaccca gggggactcc 53340
acccccctaa gcaagcacac acatactcac agttgatgag ttgctggtct ttgagtccca 53400
gctctcttac cctcccttta ctccaccagc ccgacgaccc atgactgagg aggggatttc 53460
tacagtctca ggatttagaa agtctgtaag ccatccatgc tccagaaagc accgatctgt 53520
tgtagttgca aaaacaactc tgtaatttgt tgaggttctc aaactgacag ccagcgagac 53580
tgggtgggag gccctggatc tgttctccct gactgcggga ggagcagcca ctaggacttt 53640
agcaggaagc ccacatggag gctccgcagg ctgtggccca gctggtgatg gccctttgc 53700
tcctggcagc ctgaggcaca gctgcctgta ttgtcctcat ctgttctgac tgaaggatgg 53760
aggtgctgaa taaattaggc ctcaggcctc taccaccaga gagctggaga atgggtccac 53820
gtcattcaag gacctgaatt ttttatgctc aggagcattg gaatcctctt cttccaggga 53880
ggaattagcc tgcaaggtta ggacttgaag agggaaggta tttaataact gggcgaggat 53940
gggtgtggtg gctcacacct gtaatcccag cattttgngg aggctgaggt ggccagatcc 54000
caaggtcaga agatcgagac catcctggct aacatggtga aaccccatct ctactaaaaa 54060
tacaaaaaaa aattagccgg gggtggtggc gggtacctgt agtcctagct actcgggagg 54120
ctgaggcagg agaatggcgt gaacctggga ggtggagctt gcagtgagcc aagatcnggc 54180
cactgcactc cagcctgggc gacaggagca agctccgtct caaaaaataa aaaaaaaaaa 54240
aaaaataggt gaaaattcct tataaatcca ggattggctc tgagagaact ggctaagatt 54300
caggaagaaa caaaaaattc agaatcctac aaggttttga tgacaattag ggccaaaatt 54360
ttaggaggag atgtaggatg caggagaaaa ttaaagtgtt ttctttatat cagaggagga 54420
aatagtagag gtcagtgaag gtctggggta gggaaacatt cagactgtcc attgcatggc 54480
tgtggagtga gactgcccct agcctgggcc agccttcctg cgccacaaat tgggcatccg 54540
tgatgctagg taactgtggg aacaaaatga cagcttagag cagccatggg tgatgtttgg 54600
tggtaaaaaa cctacaggcg tttggggtcc catgattgtt ccagaccatg actcttcctg 54660
gttgtgggtt tgttacagag caggagaagc agaggttatg acagttatgc agactttccc 54720
cctccttttt ctcttttctc ttccccttgc ttttccactg tttcttcctg ctgccacctg 54780
ggccttgaat tcctgggctg tgaagacatg tagcagctgc agggtttacc acacgtggga 54840
gggcagccca gtactgtccc tctgccttcc ccactttgag aatatggcag ccccтttcat 54900
tcctggcttg gggtagggga gaccattgaa gtagaagcct caaagcagac ttttcccttt 54960
actgtgtgta ctccaggacg aagaaggaag atcatgcttg atacttagat tggttttccc 55020
agggaagagg gcggagcaga gcaaagtcac tgtgaaccct gggccaggcc ctggctgggc 55080
cagctcctga gagcgtctcg tgttgcagac ccttgcccac ttcacccacc tgcaccttct 55140
ccccctctca cagtgtcact gctgctaatg gtcaaagtca aatgtgtggc cacatgggat 55200
gggccaggtc ctctcaggct actttctgga tgtcattttt aaaatatgga aacatgcagg 55260
tgccttccca aagaggcttg gactggtata tccaacgaga aacaaataag ctaaagaaag 55320
tttaaactca agaagaaaga tgttgacagt ctatgtaaca gctggaaagt ttataggcac 55380
ccacctttgg gacaacccag tgattatgaa catgtgatat ctactattta aaagaaatgt 55440
tctcaccttg ggttgattgt ggtataccat gtgttatgaa aattgttgag ctgaagcttt 55500
gaatcgattt agttgagtct gactcacttg ctttggttcc tgtgtatttt actacccctc 55560
ttgtcagtga ccttccttcc ccaccccacc cagagtgaat ttgtagcatg attgtataaa 55620
cctctatgta gaaaatggag atttcttgct ctgaaatgtt aagctctaac tgatccattt 55680

ctgtgtcctt tagcctagta tgtctgaact tccattcttg ttatatattt aaactttccc 55740
tctatattat aggttttgtg gcatccacgg tcaggtgtag aggaagctgc cccttgcaga 55800
actgtactgt aatatttttc ttttataaat attttcacag gactgattgt acacagggct 55860
tgtaataaaa ttttaacact gtgctgtgaa acaactatgg ggaatctcca ttgaaggcta 55920
cttcatgggc acctgaaagt ggagtgttat agctatgact ttctatttct tgtttcctaa 55980
gtaaattaaa cctaattttc acccttтcat tctgtttcag cctcctgtat aagaagtacc 56040
gtattttctg cccatcatac tttgtaataa aacttgaaca tgtatagatt gactgaattt 56100
ctttttgtga cgaattcagt tcttcccatt ttgtcacttt ggtcatgttc catagacagt 56160
ggcagtgccc cagcccgagc cggncagtnc tactcatcac cttttatatc ctatttagtc 56220
attcacttgt taattcagtg gagcacctcc tctatgccag gggacatggc cctcaagttg 56280
gttgcagtct accagaggag acatctacat aaatacttgt ggccgggcac agtggtctgt 56340
gatcccagct gaggcatggg ggtggctgag gtgggaggat cacttgaggc caggagttca 56400
agaccagcct tggcaacata gcaagaccct gtctcaatca atcagtcatt tttattaaaa 56460
aaaaaaaaat ttcttttagc cagactaaga tctaagacaa atctttttтt ttgcatttct 56520
tttttctttt tttgagacgg ggtctcgctc tgtcacccag gctggagtgc aatggcgcga 56580
tctcggctca ctgcaacctc cacctcccgg ttcaagctat tctcctgcct cagcctcctg 56640
agtagctgag attacaggca tgtgccacaa cacccgactg atttttgtat tgttagtaaa 56700
gacggggttt caccgtgttg gtcaggttgg tctcgaactc ctgaccttgt gatacgcctg 56760
cctcagcttc ccaaagtgtt gggatggtac cgagctcgaa ttcgtaatcg tggtcatagc 56820
tgtttcctgt gtgaaattgt tatccgctca caattccaca ccacatacga accggaagca 56880
taaagtgtng agcctggggt gcctaatgag tgagctaact cacattaatt gcgttgcgct 56940
cactgnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 57000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnntattc tagtcaaaaa 57060
caactggcat gcagcagtga tacctactcc ctcttatgat ggccaaaaag aatggcattt 57120
tgactgttga cattttaggt gttttgccca tagggtatag gcccnaaatc ttnttgncct 57180
ggaaaactga gagaccctgg ggagaagcgg catcagcggc atatttgtaa agctacagga 57240
cctgctcatt cgcctggaca ggtgagtgcg ggtggtgtgt ggcgttccaa atgagcagga 57300
gcactttcct gacagacact gagaaattct tcaaaatgca gtggctcctc agcgagaggc 57360
ggttttaagg agtaagatgt aaagatgcca gcagtgctga gacctggcca aaaagcaatc 57420
tgagaggcca ctgggctcct ggctagaggt ccccgctaga gctacacatt tcatcttcct 57480
gatgtgagag agaggggtat agagatggag agaggacgct gatgcctacc ctggtctagc 57540
tgagaaggcg tcgctgatgc ccaccccggt ctagctgaga aagcagtcct tagctatcag 57600
ggagtgtgga aagagactgg ggcttgagtt agtcttgcta tttgctttcc agccaccaga 57660
gacaagtcct gccctgtttt aggtttgggg gtggagaggt ccgtctattt cgcagaatct 57720
aaattagtaa agaaatgctt ctgatctcag gttcatgctc cctgtgtctt aatgatccct 57780
tcaaagggca cttacttact tacttactta ccagacacat gggactggtc gcagccagat 57840
ttttattcag gaaactgccc ttcctgcagc ccttccctaa cagagcagaa tgattcctaa 57900
caacagagaa cgaatgctcc tttgtttcgt ttcagttgct tagttgagtt gtttagttgt 57960
aagagctgat gtacagccag tgccaggcct gtgccgaggc tttggttcca atgttcttgc 58020
tatccactca atagtggcct cgctcactca aaggtcagcc taccaaatga ccacgagacc 58080
tagaagatgt gtcttctaga caagcttcaa gatcagaagc ttgctgtcta tctcatgccc 58140
ccacccaggg ctgcaaaaca gggctgaaag gggtctttct tcaccctgaa acagtttttg 58200
ctgcgtctga ctttaaaata tcacactctc ataaacactc ctagaaaaag aaaagaaaca 58260
gttcctgtaa taatagataa tcaaactata gatctcaaaa tggaagtaga caacaacggg 58320
catgaccctg ggtggaaata ttgggactca gtctcatggt cccaagagat tccaatctgc 58380
ttctgaaaga aacaatttct gtatgacacc atctgggttt acaggaagag ggccactaga 58440
acaaaatctg gctgtaggtg aagatccaac taacgggatt gtcacaccag ccagccttcg 58500
gttagagttt ccttttagtg gagtatgaca ctcagtgatg tatcaggatc taaaaagtag 58560
atgtctccag aacagacttc ctatagcctt tcactgagaa tccagaacaa tctttttтgt 58620
ttttctatgt attcattccc tcccccaccc tggaaaagca gcagctttta gtggttaatt 58680
gtcggctaaa aatctagaga ataactctca gtgaattctc ttaatagcca caattctcat 58740
ctaatttcct tggctgactg caatatttaa agggctggct gtttttgaca tgttcctaca 58800
atgcctcttg gaaaatcaat gtagtgacat tttacataat tatattgggg ttttcttcca 58860
atttagcact taattcttct ctctctgatc cttttacccc ctaggaagac cagcaataat 58920
cctaatgcac actgcagacc tggtggctgc agggcgctgg cggtctgggg aggggccaca 58980
gcatgagaat gcgctgcggc tggctgtgtg ggtgacactg gctaaggatg caaaggcaaa 59040
ggggcatcgc actctttcta aagtttctag aagaaaacga acaagaacac cttatttttt 59100
aaaaaaggaa aaagacaatt acacaacaag aacatcagtg aaagcgattg tctcctggaa 59160
aaagctgacc agtgtgtctg atctccctgg gttaaagcac acagcacgaa aatctcctcg 59220
ttggcatctg aagagaggga gagagcatgt ggctgggcac agacaagaga aactgaagtc 59280
ctggctcccg ggaggccgct tcccagtcgg ctgcatgcaa aggatttcag catgtgggct 59340

gccacctctg aacaccacac agaaactaca tgagaaatta agccagagaa ctacctcgaa 59400
tgtggaaacc aagcctgaaa atgtgcagtg aaaacatatg ggtctcactc caacaaaacg 59460
tcttttaaac attagactcc acgaacggtc tctttggttt gataggggat gtctgccctt 59520
gccgagagtc tcgggacagc tgcctgtaca ggttgtcctg gtatgcctga gccacaggta 59580
gagtccgagc taccttcacg tcgggatagg aggaggcctg gctgactcgc tccaagaggt 59640
ctccacgaga cccgttagcc agcatcccat gggggctgta ataatcgtcc tccagcaaat 59700
ggccattctg tgcattgttg gttttgttat aaaaggcaat gttgctgatg gatgatggtg 59760
gactgaagga ctcaggctga ggcaggttgc cnggagacgt gggactgagg acggtgtcca 59820
cagatgacac tgcccaggtc cgattctgct ggtggaggtg ggggtactgc tgagtccgtg 59880
ctgttttatc tatgatcccc atgaatggtg tggtcctgga tcgggtgggc tcactggggt 59940
aacctccttg agnttgngag acactggcga cagctcngtc acatgacctc tcatatgcag 60000
gtttcagtgg gatctccatt tgctggatgg aagaagatgg ccggaaggtc ggagttaggt 60060
tggaggtgga ggagatacta ttgctagatg gagagaagcg aaggcctaca ctttgggaat 60120
gcagcaacgg cctgggagtt ggcgggtgcg gtttgatttc gttagggttg acactgatag 60180
gtcttctgat taaatgcgac acatcagggg aacctagttg gctggaggag cgctccagat 60240
ctagttttga gtgacagact gggtaatagg atgctgactg gctgcgtccg atctgtggtg 60300
tctggctgta tctcacgcca cctcggtatg ctgaggaagg tcgctgtgga agatcctctt 60360
tggtcaatcc tccatgctga cagatggcgc ctgcactcag gctggcttgg acatgctgca 60420
ccggccttcc atcccctacg attcccccaa ttgaccctg ataaaccaac cggctctggc 60480
tgactcctat gtctccttgt gaagactggt atttactggc cattgaatgg gctacttggc 60540
cataggctcc tgtggggatg acggtgcttg aatggacggc tgggctgggc tggttacttc 60600
tcacaatctg ggccttggca ggctgtagcc tgagcccttg ctggggaact gccacaggga 60660
gctgaggcat ctggtagggc cgctgggcca ttttnggata atggagattt gggtctgcca 60720
ggaagctgac tgacgcttgt ttcctgctga tagcgcacgg gtgaaccaga ctgggatcta 60780
tagacactca tactttcagc tggagggctg gcccgatact gagggcctct ccggagaggg 60840
gaagggggag ggcttgggta ttctttgccc tgtcctccca caggaggggg gctgaaacgg 60900
taagatcctc cctgatgagc cggggaagtg tgtggtgggc taggcctgta atgtgagttc 60960
tgatgagttg gagaaagggc aggtgaggtg gactgatagg tctgtctagt gggagagcct 61020
acagaactac tggaccggaa gtcaaaaacc tgatgagagc caagaggtga gctggagatg 61080
taggctgagt cccgatgcgg gggagaggag ggtgggctct ggatgaccgg gagcccctgg 61140
ctgggccggg cctctgtctg gaggccanat ggaaacattt tcaancatcc tgtntccagg 61200
tactcctcct cgaatatatc ctgtacagag tatatgtctt catgctgtgg ctcaggttct 61260
gcttcttccg gcaactgctg ctgaggctgc ggtggcggcg gtggctgctg tggctgctgc 61320
atctgtctac actcttcttc cactctcagc agaagtctct ggatctcacg gttgttggga 61380
cacagcttga tggcctcgtt caggtcctct aaggctgctg cgaactgtct gcttgacaaa 61440
atgtgaacac aggaactgga gttattctga atccagagaa atcaccaaga gctgagaaca 61500
tggtcactat ctcagagagc tgtggtggcg tctgtgggag tctggggtat ttcggtttag 61560
ttggatttat agggaaggaa acaggaagac tcattctgga agagtccgaa tgattcccta 61620
acccattcgg cctccatctt ccttctggcc tctctgggaa tagtggagaa atgccccatt 61680
ttgtcagtcc angccaggcc cttcgggaag ttgttttggt gcccaaacac acaccaggca 61740
gtatgagagt tgggtggggt tgtcatttac agggtcactg tcaagtagct taggaacaaa 61800
aaaaactaca tatttaaaaa tcttgctgat ccagcagaaa tttttcttta tattctcaat 61860
attttagctc caaatttcag gcctgttcta ctctaggcaa gaaccctgtc ctctgccctt 61920
tctaaccctc atctctcaga attgctgtgt ctccatttgc tctatcaaag tggggcaggc 61980
attaggcttg ctgactaggt cactcggggc tggagctgag ctgactccaa ggagactcta 62040
atttacaaat gaagaaatga actgtcagag acattcccat cttctctact aatatcacct 62100
gacaggggca actgaccttg tgaaaaagga actnacccag aggnaangga gaatttcagt 62160
cagatttctg tgaggacttc gttgccccag ctaggtaaga gtgcgatcac tgccaccacc 62220
atcactagcc tcaagctggg tgtgattttc gggaccatgt gcccctggcc atagcagtga 62280
acctgaagca gttctttccc atgcagctgg ggaggaggag ggcaggaggc ataagagaca 62340
gtgacaagtt acacaaagat cactgatggc aggttctagg ttattctggg agcttttctg 62400
cccagtggcc aaactaaagc agatgggcac aaggaatgtg gaatggtgtt ctcttagagg 62460
gtcctctcca tcctggcagc aaggacctct tgactctgtg tgtgtgtggt gggtgatggt 62520
ttaggtggaa gggggtctga tgccaccaga cttgaggaga gccccagagg gaagactaac 62580
taatccggga ttcctcaaac gcctatggga gaaggaagag gaacctgtga tgttaagtga 62640
accatcctac tttccctcta aagcgggggt ggggtgctct gcttaagtga tttctttaag 62700
tcgccgttgt ctctttgtgc tagctgctct gggggttgaa aaatgcctga gatacggcag 62760
aggcttcttg gaagaagctg ccattctcca gtgaaaaagt gtttttctcc cgaagtgaaa 62820
cattcgtggc ctccacagcc cagcagattt gctgatcttg agctgagggc tgtgctatgg 62880
gagccatccc caaagcttgg ccctgacagt tctcaaggac agagatgtat ttgggtacat 62940
atgccccagg gtggccgacg ctgccttggc caaattttca gaaaagagac ctcttgcttt 63000

caccctctct ctctctcctc acctgctgct gcgttttgcc ncttgntctc gcatagtaag 63060
cttcataaga tttcggtttc agctccaggg ccttagtagc aaattcctcc gccattccaa 63120
aatcctgtca ttacaatagg tgtgcaaatg agtcattgat gagatataaa aaatagacaa 63180
cttgggaata aaacaaatta tttgtacatt ttttgacatt catggatgga acgcctggga 63240
cgctagacgt gcctgacatg gctcaagccg cgtggcctgt ggactgaaag ttccttctgc 63300
ctttgtcctg tgtagcgctg cgactgctca ctggacttta tctgcccaga atattaaaat 63360
agttttaatt ttccctcttt ttgcctttct aataaagttc tcagcccgga agccaaacac 63420
caacccttc cagtgaagct cctgaagtta tccatgcatg taggaggcaa acgcacacat 63480
gtccacattt cacgtggacc ctggtctgag ccccacacct tcctgctaca tgcactacac 63540
tgtggctttc taagggacag gtgagaaggt gtcctgtaaa gtccgggagg atgacgtctg 63600
taaaggaaac tgcactgact acagagacta gagggagagg gggtgtcaga cagacctggg 63660
ttcaattctc tgaccagctt cataaccccc aagtgtatag tgtattttct catccatacc 63720
taccccgggg gtaggatcaa gtgtgataca gtatttaaaa gccctgcaca gttcccggca 63780
catagtaggt gctcattaat tatcgtnttt ctttctattt cacagaggct tctctgatct 63840
gttccttttt ttaactggtt ccttanctta ggaaacactt aaaattaaaa ataaaacaaa 63900
acaagagaag accccagctg tgactgaggt ctgtgctctc cctagaggaa cacgctcaca 63960
ggcatgtcat tcctgagaac ccctggaaac agagactggg gggtggggtg gggaacaacg 64020
ccgactaggc agncctatgg ttttcagtat tcccttaaat actaagctaa aagagaaggc 64080
tttccccttc ttttagtgtc aaagccacag tatagcatgg tagcctctgg ttatggtcat 64140
gagtcctaag gcagaaaaca aggccttggg cagaacaaag gccctggggc ataggtaaaa 64200
aaatttttta aaagttgatg ggaaatcctc tacctaagaa gacagacgtc tggtgtttgg 64260
cacagtattg gcccaggagg tgattaacaa ccgcttctga acaaatgtgg agctagaaag 64320
aatttcagct tttatgggtc accctctccc tgctctgctt ccagccttac ccttttattt 64380
gttcttttgg gaccatcatt attcctaaaa atatgagggc atttttaaaa agtctattat 64440
ttggtaggga aaagagatca gtgggttggc agcaacttct gagctttgcc tcatcctcat 64500
cggctggtgc aggggaggcc ttggtgctgt tcctttcatt tcacaggaga ccccaagagc 64560
atgaaaatct gtctcccgag accgccacct tctcccttcc cggggttggg gggcctcttc 64620
cacatcccca cttccctcag caccctggag caactctgcc tctgagggcc atggcgttgt 64680
tctctgagac agtcaggttt aactctgcct ctgaaggtca tggctgtgtt ctctgagata 64740
ggtttaactc tgcctctgaa ggtcatggct gtgttctctg agacagtcag gtttaactct 64800
gcctctgaag gtcatggcta tgttctctga gacagtcagg tttaactcag gttccttcca 64860
gtgtcagggt ttagggtctg gaatggatga tctaaaggtt acacacacaa atctttctgg 64920
aagatccctg gccatgaggt tcaaatcttg cttctgtcca gctgagtttt taattgatca 64980
atacatctgt ttggcttgga ttatttttat ctcctgagga actgtgagac ctcatgcagg 65040
tagatgtcag agactcctga cacaatctta gctatttcca aaacctgaga gagggttctt 65100
ggcttggagt gtcttcacgt taaagggcag tggctgaggg agcagtgtct ggggacatgt 65160
gagatgggtc tgcaaggtag gcaggaagga ccgtggctta ccagggaaac aataaaggtn 65220
cagggggnat ttctgacatt ccccctagga caagaaggga catagcatgc tgggatacaa 65280
aaatgaggaa agaagcgaaa agttcgatag cctcaaaaat taaaccttca aaaccatggt 65340
ccacccccctc tccttacccc ttcttcaaat gataagggaa tatacctgaa attaatgact 65400
acaattcaca tctaattgca actggaattc tagtgaccac tgctctcaca tttccccnaa 65460
agaccagaac actgacttgg tggtatgggg gtgggaaaga tggtattctc atcactttca 65520
gttaaattat gaacattcag aaagctgcaa agtactcgca gcattaattt ctgtacagtt 65580
tgagatacag attgtactta gacttgattt ctttcaaatt gcaaatgtca gtgctctgtg 65640
aaggacaacc aagttccaaa tggggcagag gaggatttca ggggcctggt aaaagctagt 65700
gagtgacaca ggctatcagc taccaggatt gcagaaagga gttggggtga cagggactta 65760
cgttcatttt cctgcgacac cgagagaggt tgaggaggag agacaccttt agttcccgga 65820
aagttttcaa gtcctcacca aacccttctc tagggaactt cttcagggcg tactggtagc 65880
gctgggcagc ttcctttact ttacctttct aatgggagca caagatggcg aaacacacag 65940
ttggtctgct gaacagctga ggaagtgtgg tccccactcc cttcttcttc tcagcagctg 66000
ttactaaggg ctaaccctcc caaacgccag cctagagagg agttctggcc ctgatggtga 66060
gggcagaagg gtctagtctt ttctatagga aactcacctg gcaggaaagg caaggttgag 66120
gggataccca agatacaagc cagggagcaa agccttggca aatggaaggc cctggaccct 66180
gacttcccaa ctggattcca aagaggctga ccatgctcat ttcaaagttc ctttgctctc 66240
tataacacat atgacttcct aaccccaagc actccaaagt tattacttcc tccttaattt 66300
ctggcttgtc tggagaactg ttttttggt ctgacgctcc acctggccag ccctctagaa 66360
gtcactctct tcctcatgaa ggatgtggga gctaaagcag tctagcagtt gcttccctaa 66420
tctcatggca taaaataata ccaatcagtg ttccccattt ttggaaaggc atggagtagg 66480
gtggcagtag aaaagtgaag tttagagttt gtgattccat ctgatgttgc aggagggatt 66540
atccaagagg aagtaaatct atgaccaatt tagagatctg gccacagcga ccttgaactt 66600
ttcttccact ttgtatctga aaatctataa gctgagacca gaaagatgca ggcggcatct 66660

cccacccctt tagctgatct ttatccactc tattccttct ctatagttac aactccacag 66720
ggatcccagt cttttacttt tcccttagca ccattcacta aggagagatt tggagtcttt 66780
caaggtagaa ttctgagggg aacaataatg gcactgcata ttttcattaa gaggactcct 66840
cgcctcagaa taccagaagt tatgcnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 66900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 66960
nnnnnttttt ttttttttct ttttttttct tttttgtatt ttagtagaga cggggtttca 67020
ccgtgttacc caggctggtc gcgaactcct gagctcaggt aatccgcccg cctcggcctc 67080
ccaaagtcct gggattacag gcgtgacacc gcgcccggcc agttttactc ttcttaaatc 67140
tgatgttcat ggcataaagc aggagtagtg gctttaaata tcactagcca cagctttcag 67200
ttgctctgat cgctatcatt cacaactttg agtaaaagga agaaagaaaa ttcttccaac 67260
tgacttattc aactttcaaa aatgagtttg ttgcaatgcc agtttaaaag ctcccctact 67320
acaccgtgct tctataaagg attctcaaat tgatgtatgt ccagcaagac ttactactaa 67380
aggtcacaca actgtactag cacgaaacaa gccaagagat aaggccctgt catccttctg 67440
tgttcaggaa acaacggcgg gatcatgcca caggccaaga tttccaaaga gattgcttta 67500
gagagaatgg atatcaggaa gcaagagcaa cagtaattct ggcagaaggg ctaggatcca 67560
tgagtacccg gggggggtng ggagtagggg gcggggaagc ttgtacaata gatcgtctta 67620
atagaaatct agtgtggcca ggagtgcacg accaccaata gaagcgcctg cattttcctt 67680
cacagctatc actctctctt cccttccccg cccaccaacg aaagcaatgc ggcaggagtg 67740
gagagcacgg gatgaataaa tggagtatct ttggggcgta aacccggttt taagacctgt 67800
ttccaccccg aaatagctgc ctgacccagg caagtcacaa aaccgtgcct cagtttcccc 67860
actacactgc cgactacaca agaatgctgc ctggattaaa ggagacagac aacgagtacg 67920
aaaccttcag cacagggcca gacagaggaa ccgctccgtg gccctagggc gagtccgcaa 67980
gagggttctg aggcctgggc tccggcccgg aaangggcgc tcccggggcc cgctccccgc 68040
cagctgggtt ccgagcccct gcccggccca ccttgttgtt gtactcctcc acgaagctgc 68100
ccagcgccaa gcgaaagcgc ttnatcgggc cgcacactcc agttcatcgc gtagactgtc 68160
cagggcgctt catacttgta gatctccttc cgtttgccgt gcagggacat ggtggccacg 68220
gggncgcagt nacgggccgg gtcaacagtg ggctgcggcg ggcggggagn cgagcctgag 68280
atctatgggt ccgaagggag gagaggaggg ggcgggcagg aacggcctaa cccggaagcg 68340
gggatgcggc gacaaacaac gacgacggcc gagcccgacc cctagtttca aaccagcttg 68400
ggaacggacg acaaccacct tccgttttgg gacgccgccc cgccctctcg gaacggaagc 68460
cgccggaccc ccgcagcggc accggccgtt ggttgcctga cacgtccttt cgaaaggatt 68520
ctttcttgtc attggttatt gcggccgtag cgtctgaaat ctcctcgcta ttgggttaaa 68580
tgtttgtcat tgcaactact cgccccgccc aactctcgga ggcaggcgct ttggcagccc 68640
caaaggtcat tggttggcca agatgtcagt caggcagata acggctcagt gcgggtggtg 68700
ggggcgtggg tggcctggag gcgagcgtgc tgtagcagcg ggcctccaag ttctaggcca 68760
agtctctgag agtgaaaccg tctgtgacct gcgctgactt cccctccgct gctgctgttc 68820
tgagcggcct ctccacgctg tcgagtaaaa gtacaattct gccttagtgg aagacctact 68880
gactttcgcg ggacagagcc ctgcggcctg gcagncctgg cctgcgccca acgggtacca 68940
ccttcccgcc ccatctcttc ccagggccct tttaacccga aagctgctct ctccgtgcct 69000
cagggacagg cttcgggcgc gcagatgcgc tcagggccag gccagcttga gtcaggccct 69060
cccgcgcttt cctgcaggat ctagaaatgt ccccgaattc tggggtaggc accgacccga 69120
ccagcttggc tctgtttttg tcagttccca tcctgtaccc ttcccgccgt ggatcccaac 69180
gcaaatgcta atccagcccc aaaggtatct gtgcctgtga gttggaaaag gggaagggcc 69240
agagaccaaa tgacaccttt atataaatta tttccaattc tcagctttca tcaaaattga 69300
aggaggtttg tattgaattg ttggatgata acatntaaat actttttttt ttttttttna 69360
aagacagggt cttgccctgt cacccaggct ggagtgcact ggtgtgataa gggctcactg 69420
tagcctggac tgcctaggtt caaacgattc ccccacttta gcctctggag tagctggaac 69480
cacagacatt tgccaccaca cctagctatt tttaaattat tatttgtaga gatgagatct 69540
ctccatgttg cccaggctga tctcgaactc ctgggctcaa gcgatcctcc cacatcagca 69600
tcctaaagta ctgggattac aggcttgagc caccattctg ggcactaaat ataattttta 69660
atgataaatg atttttccac ctcccaggtt caagtgattc tcctgcttca gcctcccgag 69720
tagctgagat tacaggggtg caccaccaca gccagctaat ttttgtattt ttagtagaga 69780
tggggtttca ccatgttggc cattctggtg tcgaactcct gacctcaaat gatcttccag 69840
cctcggcctc ccaaagtgtt gggattacag gcgtgagcca ccgcatctgg ccaagaaatc 69900
ctcttttttt tttaatttta ttttatttt gagatggagt ctcactctgt cgcccgggct 69960
gtagtgcagt ggcactatct cggctcactg caacctctgc ctttagggtt caagtgattc 70020
tcttgcttca gcctcctgag tagctgggat tacaggcgcc cgccaccatg cccagctaaa 70080
ggaaatcccc tgtaaggagg aaagtttaaa aacaaattta aaaaggaaag aaaaagacgt 70140
acaaaattcg cggtacctgt cttattctaa taaaaaatat gatatttatc cacagatata 70200
ggaactaacg gggggagaaa attttctctt ttatgacgta tatttccaaa aaaaatttta 70260
gtttttatta gaaaattgaa tatatattta tgttgttttg aacaactgta ctactaatag 70320

atgtaaccct taaagcctta ttaaggttac aaaaattaat tttaaaatgt caattaattt 70380
atacacattc gttttgaggc agggtcttgc tctgttatcc aggtggtggt gcagtggggc 70440
aaacagtgct tattgcagcc ttgacctccc gggctcaagc aatcctcctg cctcctttat 70500
tttttgtaga gacaagatct tgctatgttg cccaggctgg gttcaactgg tcctcccacc 70560
tcagcctccc aaaatgctgg gattataggc atgagccacc ccatgaccga cctacttttt 70620
ttttgaaacg gagtcttgct ctgttgccca cgctggagtg cagtggcacg atcttggctc 70680
actacaagct ccacctcctg ggttcacgcc attctcctgc ctcagcctcc cgagtagctg 70740
ggactacagc acctgccacc acgcctggct atnnnnnnnn nnnnnnnnnn nnnnnnnnnn 70800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 70860
nnnnnnnnnn nntgggttgt nttttcattt tctttctttt ttttttgag agggagtttt 70920
gctnttgttg cccaggcagg agtgaagtgg aacaatttgg gttnaaaaaa atttttgcct 70980
ccccggttca aggcaatccc ctcccccacc cccccgagta gctgggataa aaggaatgaa 71040
ccacaacgcc tggctaattt ttgtattttt agtagagatg gggtttctcc atattggcca 71100
gcctggtctc gaactcccga cctcaggcga tctacccacc tcggcttccc aaagtgctgg 71160
gattacaggt gtgagccact gtgcctggcc gtcttttcat tttcttgatg gtgtcattga 71220
agcacaaaag ttttaaattt tgatgaagac caatttatct gttttttctt tcatcactta 71280
tgcttttggt gtcatatcta agaaaccatt gactaatcca aggtcacaaa gatttattgc 71340
ctatgttttc ttctaaaagt tttatgattt tagttcttac atcaaggtct attttaagtc 71400
ttttgttttg ttttttgttt tttgttttga gacagggtct tactctgtca cccaggctgg 71460
agtgcagagg cacaatcatg gctcactgca gcctcaacct cttggcctca agcaatcctc 71520
ccacttcagc ctcccaagta gctggtatta tagacatgcg caaccatgcc cagctaattt 71580
ttttgtagag atagggtttc accatattgc ccagactggt ctcaaactcc taagctccag 71640
tgatccgccc acctcagctt cccaaagttc tgggattata ggcatgagcc actgcaccca 71700
gccccaaatt ttgtatatgg tattagaaag gggtccaact tcattctttt acatgtggaa 71760
atccaattgc cccagcacca tttgttaaaa atattttctt tcccatttaa ttgtcctagt 71820
gttcttgtca aaaacaattg aacataattg tatgggttca tttctggact ctcaattcta 71880
ttccattgtt gagcatattt ttaagggctg tttttctctc ctgtggtaac tggtgacctg 71940
tacttcctgg aagagagatg aaaagattcc caagccaact gagttacctc acgtgggtca 72000
ggtctctgtg gctctctgca ctggcctatt cataatgata tcctcctgca ggatttgagc 72060
cttctccttt gttgtgacgg cagccgagga ggtggctgac tgcccagaca gccttatctc 72120
tttcctacct ttcaaggtta ttgtaaatac caaattagat tgtttataca caagaattta 72180
gcactaaagc actatacaaa tgtaagctat ttatttctat ttatccttct ccttcatgaa 72240
taagacccta aaaatagaag atatttttaa tttttactca ctgggctcaa ggttgcagtg 72300
tctgtattat tgcaaattcc aaattaatga agtctggctc ttcttatatt attcctgcaa 72360
aaggctgtgt gctcaccccc cggagtgtga atacgagtgt gggtcatttc ctctttcctc 72420
tgcacccttc cttcgatgag gttttgccct gtgctaggca ccatgctaaa ctctgagaaa 72480
accacagaga acaaggcaga cgccatccct gccctccagt aacatacaat ttagtgatga 72540
agctgggggа tttaacaagc cattctaata aagcgttaca gtgagggagg tgcagggtga 72600
tgcctccagc agccctggtg tcagctgctc caggtccagg ggaagacttc cattatcttc 72660
caaccctgtc ggaagtggag gcggaggctg tttattgctg acacagtncc ctaacccagg 72720
gtctatagac attgtggaaa tgccttggag tcagacggga gaatgaacca gcagaagcaa 72780
tgcccgccct ccaccctcct gaagagggtt ctcaggaact ctttggaggc gaggccagcc 72840
tctggntgna gnggcctctg gatacaggtt aggcctcagg ctcttctcct ctctactcat 72900
ctctcctccc ttnggccccc ccttcagagg ctgacagagc cccactctca tctcttcccc 72960
acccaagcct ctttccacag aaagactgct tcctcccagg agacagcagc tcatttgcac 73020
acagacaccc acagccctca aagcctggaa ggccaagctg ttaggacccc tgagagcagg 73080
gtggctcctg ggaggagagc ccaggccacc accttgccct ccctggcccc tggccttncg 73140
atggggctgc tctgatcaca aacgtccacc aacgtagccg gcccagaagt gcacccatgt 73200
cctctggtat ccactggctc tccaagccaa actgggcagg gaggagttgt gagggaaaac 73260
tgcaggtcag gagggaggct ggcaaagcgg gccagggcca ggcctgaccc cagctctcct 73320
ctcccggccc cactgccggc cagtgtttaa caaggccctg ccttctccct ctagtgctag 73380
ggacagccac cttcttcctc tccccaccgc cccctctccc ctgcaacacg tcatctgaca 73440
agtcagtgcg atctcactgg aggtgcatct cacaggaacg cggggtcaca gcctcctgca 73500
cacactccat gctgcacagc aaggtgcacg tgtccctcag agccccagac accatccccc 73560
actcacccag aagcccaagt gattcccaac agcccccagc agcctaatgg gttggggtct 73620
tgggagcagc tgtccctggc tccttccctg atcccaccgc ccagcctcac cccacggttc 73680
ctccattgcc ccacctccca ctgcgccgcc gggcctctgc cagggtcaag gggcttcccc 73740
cctctggcag cagacgccat ggtgccgagg tggcctccac aaccgccctg tgcgccaata 73800
ggacaagacn tgtcctccct cccccacact tgtcactttg agggacacgt ggatgagaca 73860
ggaaaacaca ggggagtgtg gagacctgag gtgacttgga gcaagcctct caacctgagc 73920
ggcaatttct tcatctgtaa aatngagggg gttgttctca tctctgaggc tttgtgtcgc 73980

tctcaaagcc tgctagcctc gggttctagg actctgttgg gatcgtgtgt gatgttttct 74040
gctgagcgac tggcagcctg tgtcctcggg gggaaagagg gcaggcgctc caaagctcct 74100
gcgctctgtg gctccccctc cctcgcagcc ccaagcccca ggtgtgccgg ccgccctgag 74160
cccctccagc acctcccgga ggcgcctgca agacacctaa ggtccccgcc tccctcctct 74220
ccccccgcc acacccctac ccccggcagg cgacgtcccc gcccctcgac catggcctgg 74280
tgaagaagcc ggccaggccc gatcagcccc aaccccgccg cacgagcggc gnctgncgga 74340
cagctcctgg ggccccggcc ttgtcactcc ggaggcggga ggctccgggg ggtcgggctg 74400
ggaagatcga gccggaggcc gctaggctnc ccaggccccg gccganggct gcgcggccgc 74460
acggtgggca ggctcgggtg nttccggcaa actgccgggt ccccnatctt caaaagagag 74520
gaggcccttt ctccagcttc ctctgcggga gcccgaccca gccccatccc gccacccccg 74580
ggctgcacct cggccccntc ncnccggcnc ncgcgccccc gcccggggcg ggccagngaa 74640
ncctcggccc gcgccgctng gggactttgg agcggaggag gaagcgcggc ggggcggggg 74700
cgggggtgtg tcgggtttta ntaacccgca gggcggccgc ggcgcaggag aaggggcaga 74760
gccgagcacc gcgcaccgcg tcatgggggc cgcctcgggc cgccgggggc cggggctgct 74820
gctgccgctg ccgctgctgt tgctgctngc cgccgcagcc cgccctggcg ttgngnaccc 74880
cgggctgcag cccngaaact tttctngctg acgaggccgg ggcgcagctc ttcgcgcaga 74940
gctacaactc cagcgccgaa caggtgctgt tccnagagcg tggccgccag ctgggcgcac 75000
gacaccaaca tcaccgcgga gaatgncnaa ggccaggtgg cgcccgggcc cgggcggggg 75060
cggggcgggc cgcggcggcc aatcacagca cgcggccggc ttgtggggnc gggcagnntg 75120
gcgcccccgn acccgaaccc caccccgacc ccggaccctc gccccgacag tcagccgcgg 75180
ggcccgagcg ccgggctgcg cgcacggcct gcgctcccag catgcacgag ttggatggat 75240
gagggtggct gctcccaggc cgcgcccgcc ttcgccgaag gtgctgggct tggctctggg 75300
gcccccgcgc tctcgggcag ctgccttctc acctccggac gctgtcgctg tcaccgtcac 75360
cgcactgcac tgtccatcca ccctccactc gcccggcctc ttttctggtc ccaatttctg 75420
ctccaccatt cccatgaggc agattccctn ccagaaggag gaagcgcggc ttccgcnaaa 75480
ctaaggtctc ccgcagggat nctccccagg gcccgggctc tggaagccct tggccttcct 75540
cccctccccc agcaccgtgg cttctccttt atggcctgca tctanagcag ggtcctacac 75600
cctccctgcc ctcctggtgc ccaataggag gaagcagccc tgctcagcca ggagttntgc 75660
nggaggcctn ggggcagnaa ggccaangga gctgtatgaa ccgatctggc agaacttcac 75720
ggnaccncgc angctgcgca ggatcatcgg agctgtgcgc accctgggct ctgccaacnc 75780
tgcccctggc taagcggcag canggntggg ctgagggctg aggcagagct cgggggcggc 75840
ctcctagtgc cccatcgtgg gggtcggggg agagcagccc atcangggag ggaggaaccc 75900
tgggatccac atgggcccnt gacagaaggg naaagcccag gtaagcacag aatggctttc 75960
tgagncattg atttttcttg gagatggggt ggggagttac tttctgttaa aggaagcatt 76020
cntggagtag gaagccaaat tcaaatacac ttctccctag gctggtttat gagcttcttt 76080
ggaagagttg agaagggctg ggcgtggtgg ctcaagcctg taatcccagc actttgggag 76140
gctgaggtgg gcgcatcgct tgagcccagg agttcaagac cagcctggcc aacatggcaa 76200
aacctcgtct ctacaaaaaa aaaatagctg ggcttggtgg tgcgtgcacc tacagtccca 76260
gctactcttg aaactgaggg ggaaggatca cctgagccca ggaggtcaag gctacagtga 76320
gctgtgattg cactactgca ccccagcctg cgtgacagag tgagacctcc ccccaaaaaa 76380
aagagagaga gaaaaggttg agaaagactg ggaagtcacc aaagccagag aatgggaggg 76440
atctgccctc actgcagggt ggtgccaagc tgggacttga ccctgaccct gactttcagg 76500
actcctgtcc cccactccac aggctgcctc cactggcagg ggactcagaa gtgatccggt 76560
cacactaagt gacacttagt gatcagaagt gccccggtgc cactgagtgg ccttgtccaa 76620
gctacatcca ctctgtgggc tcctccttgt agcagcgagg ggagggcaga tgtcccaggg 76680
gctggtcact ggagcattcc tcccctctga ctccccagta caacgccctg ctaagcaaca 76740
tgagcaggat ctactccacc gccaaggtct gcctccccaa caagactgcc acctgctggt 76800
ccctggaccc aggtacggcc cttgcagctc ccctctcggc ggtgccctag tgttcccaca 76860
ttgccctgct gcactccgga ccatgcagtt gtgtagggtc tgtggagaca gcaggtaaac 76920
ccaaaggtgt tgccctccaa ctggggctgg acggtgcaga tanccccccac gccctgcttc 76980
tcttggcaag tggacttccg gaatctccag ctgcagcccc cacttctgtg tgtacctcgg 77040
cctctcccat cacccctagg ccttcctcct ggctgcctgg tttccccttt cgtgggtcct 77100
ctcatgttcc ccaagagccc tcaggccagg gacccctcgt ggcttcccct taaaccccgc 77160
tccagccccc tttatgagca gcttcgagga aggcactcca tccaataggc cgctaagtgt 77220
ctgtctgggt tttggccttt gggtgtcccc ttggtgtcag ccaccttagg tggtcatgtc 77280
tctggggcag gggccctgcc tggngtgttt ctgtagctcc cagccccctc ccaccaggcc 77340
tgtaggtggc ccctgtctct gggggcaccg tgatgttcag gaagctggtg ggagcagtaa 77400
ggactggtgc aggctctggt gaaggccgnt tgaagacttc aacgtggagg cctcctcacc 77460
gaccctgcct gcctgtgtct cagatctcac caacatcctg gcttcctcgc gaactacgnc 77520
catgctcctg tttgcctggg agggctggca caatgctgcg ggcatcccgc tgaaaccgct 77580
gtacgaggat ttcactgccc tcagcaatga agcctacaag caggacggtg agcaggcctc 77640

tccctgtcca ggaaccacgc caggtgtcct ctctcagctg tctccccaga gtcccagccc 77700
agagtcaggc agagcagctg gtatgacaat tccagcaggc cctgagtttc ccagaaagtg 77760
gaggtgggac cggcctgcac ccagtgtgcc tggactttgc tgctggcctg ccccacgtgg 77820
ccatcctgct gtcactcctg gccctgatgc tcctctttgc ctctgggaac ctccaggatc 77880
tgtttagctg gctgtagcta attagaaatt gtagagtggc aacccccaag ccaattttcc 77940
agctagctgc agatccacgg gcctcgagcc agtggaagag ccgacttaca gctgagaggc 78000
tgaggtccga gcctttggcc tgagctacat acctcacccc cacgccccca ggcttcacag 78060
acacgggggc ctactggcgc tcctggtaca actcccccac cttcgaggac gatctggaac 78120
acctctacca acagctagag cccctctacc tgaacctcca tgccttcgtc cgccgcgcac 78180
tgcatcgccg atacggagac agatacatca acctcagggg acccatccct gctcatctgc 78240
tgggtaagga cctggcctcg cctccacatg agtcccacgg aagtgtgggt cccgaggtag 78300
gggtggggga tgtccagggt aagggaaggt gggttgtgac cctcacatct cacatgtgtg 78360
gggcatcata ctgtttgctt cacatgcagg agaccattcg tgttcccact ttacaggtgg 78420
ggaccctgag gcttagggtc gtgagggact tagtggtcag agagctaggg gccaaaccaa 78480
aggctctggc cctgggtcca gtgggggagc catcagccta gctcatgccc naaggaaaca 78540
agcactgtgg ccctgcctca ggattgagtg gctggggcct ggcacagcca gaaatgacag 78600
tggcagcatc ttgcagcccc aggacatgtg gccctcggag gagtgtgggt gggactgatg 78660
tgtgagattt ctggccctaa gccaggcctg ncagcccttg agggccccag ggtacaggtg 78720
ccggccccag ggtgccactc agcgatgcat gaagaagnca ggcacagcca ggcagggagc 78780
caagctgtcc ccttccttcc ttatctagga gacatgtggg cccagagctg ggaaaacatc 78840
tacgacatgg tggtgccttt cccagacaag cccaacctcg atgtcaccag tactatgctg 78900
cagcaggtaa gctctgggct caagcctngg ggtggtgggg gtcgggggtg gggcgcaaaa 78960
aaagggagtc acagatgggc acaggggcgg gaaggtttcg ggtactgagc agcagcctgg 79020
tgtgtctgta ggagcagtga gctggggtcg gccccctcag tgaggtgcca gctcctccct 79080
ccaggctcca cagtggcagg atgagagcaa caacgcactt tcactcatct gctgtgggag 79140
tgagggccct gcctctggga atggtggcca cagagcagag aagctttcat gcacagggag 79200
ttgacccgag atggggaccc cagccctgtc cccaggccag ccagagtggg ctccccctga 79260
cctggctcca cacccctcct ccagggctgg aacgccacgc acatgttccg ggtggcagag 79320
gagttcttca cctccctgga gctctccccc atgcctcccg agttctggga agggtcgatg 79380
ctggagaagc cggccgacgg gcgggaagtg gtgtgccacg cctcggcttg ggacttctac 79440
aacaggaaag acttcaggtt cagacatggg aagagcacgt tctggggttc cccggttctg 79500
gggcccgggg aaaggcaggc agcccaggcg cagggaagct ggttcccagg cctgcctcta 79560
ccctacccca gcactggttg gaggctgggt ctgttccagg gctagggggt ataggaggcc 79620
tattagtcca ccttctctgg cagctttgac aaatagtcac ttctatacct tggaatggag 79680
gaagaaggcc caagtggtgg tgagccaggg cagggtaaag aatttgcttg tttctgccag 79740
gcacggtggt cacacctgta atcccagcac tttgggaggt caaggcgggt ggatcacctg 79800
aggccaggag ttcgagacca gncctggcca acatggcgaa accccgtctc tactaaaaat 79860
acaaaaataa attagccagg ggtgatggcg ggcgcctgta atcccagcta ctcaggaggc 79920
tgaggcagga gaatctcttg aacccgggag gcggaggttg cagtgagctg agattgtgcc 79980
actgcaggcc agcctgtgca aaagagtgag acgctgtctc aaaaaaaaaa aaagaaaaaa 80040
agaagttact tgtttctact gcggcttcat gccccagggc agctccctcc tcattcctgt 80100
ctttcaggtg ccaatctgcc ctgtgccctg gccctgccct gttctgtcca tccgtcactc 80160
tcaccctcgc cctctctacg ccccaggatc aagcagtgca cacgggtcac gatggaccag 80220
ctctccacag tgcaccatga gatgggccat atacagtact acctgcagta caaggatctg 80280
cccgtctccc tgcgtcgggg ggccaacccc ggcttccatg aggccattgg ggacgtgctg 80340
gcgctctcgg tctccactcc tgaacatctg cacaaaatcg gcctgctgga ccgtgtcacc 80400
aatgacacgg gtatgggagg gctgagaggc ccccacccag cctcacctaa accccgctcc 80460
accccacagc aggacctcac ttgccccact cagctctgcc cttctttctg cctcccggcc 80520
ccaggtcagg cagggttcgg gatcctccta gagcctcacg gtgcacactg cgcccagctc 80580
agcacacctg ggggtcctct tccaagcagg gcccagggtc tcgagggcca gccatacctt 80640
ctctgcatct ccctggcctc actttctgct gccccgccag cccacactct taggggaccc 80700
tcttctccct ctgacctctt ccctctcctt tcatctcatc tcccaacaga aagtgacatc 80760
aattacttgc taaaaatggc actggaaaaa attgccttcc tgccctttgg ctacttggtg 80820
gaccagtggc gctgggggggt ctttagtggg cgtacccccc cttcccgcta caacttcgac 80880
tggtggtatc ttcggtgaga ggagggatag aaaagccttc gccccagcta gccctcccca 80940
gcctcctgga cagccaggcg cctcctgccc cagccagttc tagcctctcc tctctaatga 81000
tgtcccccgc tgtgacccac cgccttctcc tttcctgcct gaaactccct cttccaggaa 81060
gtcttcccca gttcctcagg atggggaagg gttgccgggt ggaaatgcct tttctacaaa 81120
agctaaatcc atctgtttgc aacctctagg ccctaagaca atttaaccat ccttttccag 81180
aaccaagtat caggggatct gtcctcctgt tacccgaaac gaaacccact ttgatgctgg 81240
agctaagttt catgttccaa atgtgacacc atacatcagg tattagcgcc cccaccccac 81300

ccacccccag tactgtcaca ccctcaatcc acttctcctc ctgtgttcct agctgcctca 81360
tccccagggc ttgtcctcat gctcctccag acctcaaagg cctggagtta gagtggccca 81420
ctctcctgag cctgtcttgg gtctcccttc tcccccaaga tagcttctgg tccagcctct 81480
gccctgcagg aagctggatg gtgcctgggt aaggaacccc tgttcctggc cccccatgat 81540
cttccctgac tcccaccctg tgcctgcagg tactttgtga gttttgtcct gcagttccag 81600
ttccatgaag ccctgtgcaa ggaggcaggc tatgagggcc cactgcacca gtgtgacatc 81660
taccggtcca ccaaggcagg ggccaagctc cggtgtgtgg tgggaagccg ggggaagtgg 81720
gaggcagaga ggagcggctg gcaaagggtg tggcaggagg tgtctggctg ctctgatggg 81780
gtggggggca ccaaccacag agctggactg atgtggatgc ctgtctcctc gctatgtcat 81840
caaatattta ttgagtgggc cttctggctg gcantggggc gacacaaatg ccccctgcca 81900
ccatcagaga gatcccaggc cccagggtct tattgccaca gtttctgcag tccattgngg 81960
gggcggaagt ggccaggggc atgtgggccg gggtccagga gcagactcca gcctgagtcc 82020
cctgtgccca tggtacccac tctgcccacc aggaaggtgc tgcaggctgg ctcctccagg 82080
ccctggcagg aggtgctgaa ggacatggtc ggcttagatg ccctggatgc ccagccgctg 82140
ctcaagtact tccagccagt cacccagtgg ctgcaggagc agaaccagca gaacggcgag 82200
gtcctgggct ggcccgagta ccagtggcac ccgccgttgc ctgacaacta cccggagggc 82260
atnaggtaaa gccctgagtg aggatggtgt ggggctaagg tgggtcctca anctctgggc 82320
ttggcccagg nccccaggtt cctggtcagc tcctaccagc tgagccctgg taccctgtcc 82380
tggagggcca ggcagccccc caagctcatc agcagggcct gcgagtgggg acaggcatgt 82440
ctttccccca gcatcctaga gagggtgtgc tcagacctga gggcccctcc ccttccagag 82500
gaagccagac acaaggctct gtgaggtcac nactgcggcc tccgctcttn nnattggcca 82560
ggggacggta gctgcaggac tctgctctcc tgcggccatg ggccagggnt tgggctactg 82620
caggacttcc cagcctcctc ttcctgctgc tctgctacgg gcacccntct gctggtcccc 82680
agccaggagg catcccaaca ggtgacagtc acccatggga caagcagcca ggcaacaacc 82740
agcagccaga caaccaccca ccaggcgacg gcccaccaga catcagccca gagcccaagt 82800
gggaccatgc aggggagggg cagggtgcca ggggtgggag aggcggggcc gggntaggga 82860
cagggcaggg tacaagggag tgcgagaggg ataatggctt ctggtgagac cacaaacctg 82920
gagaggggag gcagaggttt gtctgtttcc ctgcactctg tcccacagac ctggtgactg 82980
atgaggctga ggccagcaag tttgtggagg aatatgaccg gacatcccag gtggtgtgga 83040
acgagtatgc cgaggccaac tggaactaca acaccaacat caccacagag accagcaaga 83100
ttctggtggg agccacctcc ccacccccaa acctgagcat gtgcatacac acagagatgc 83160
tgtcccgctc accacacagt ggggctgcca ccacatttta aattgaatat ttaaaacaat 83220
actcaatttc gggccgggcg cgggtgntca cgcctgtaat cccagcactt tgggaggggg 83280
aggcgggcgg atcacgaggt cagnnnatca agaccatcct ggctaacacg gtgaaacccc 83340
atctctagta aaaatacaaa aattagccgg gtgtggtggc gagcacctgt agtcccagta 83400
ctcaggaggc tgaggcagga gaatggcatg aacccgggag gcagagcttg cagtgagccg 83460
agatggcacc actgcactcc agcctgggcg acagagcgag actccatcaa aaaaaaaaaa 83520
aaaaaaaact caatttcaga ttttgatgaa catttactca atgcctgagc aattcttctt 83580
tccttaaaaa tcagtctctg ggaggcctag gtgggaggat cacttgaagc caggagttgg 83640
agactagcct gggcaacata gcaagatccc atctctattc aaacaaacaa ataaacaaaa 83700
atcaatctct agtaacagaa taatttgtac ataaataagt ggtgctcaag tcgtttttta 83760
aaagattgaa agcctctgtt tgtctcctct acaaaagggg ctacacttcc tctttaccct 83820
cattccctgc ctatttggct gagcacaaat tatgccactg agccacacac tgttactgtt 83880
ccttggcact ttgatctgtt gcctcatctt tttctcaaca gccttgcaaa attggtgagc 83940
ttattcccat tttacagatg ggatttgata ttaactctga ggttcagaaa ggccacagag 84000
ctaataccaa gctggctcct tcnctaaggg cctttaagac acttgggggt cttctcttct 84060
ctgcccctgc ctggatatgt gttgcttgac cgcaggcatc cagggagggt gagtactgca 84120
tccaggacgt tatcagcgtc cagcttgcag agagtcttat aggcaaaggt tgcaacttaa 84180
ttccactgcc ccctcaccac cacctccagc cctcagctcc cacttggggc ctcccgctca 84240
gaggctgctc tggagctcct gggccctgtg acaccatccc cctgtgccct cagctgcaga 84300
agaacatgca aatagccaac cacaccctga agtacggcac ccaggccagg aagtttgatg 84360
tgaaccagtt gcagaacacc actatcaagc ggatcataaa gaaggttcag gacctagaac 84420
gggcagcact gcctgcccag gagctggagg aggtgtgtgg ctcgcaaggt acagggagag 84480
gggaatcctg gggcagtgag cccaacacag ggtctggcct ggccttcacg ctgcttcctc 84540
ttcctcgttg tatcaagtca tggcatctgc catgcgatng tgcacctcag aactgctgag 84600
agggcagcgc tccccagctc cctggctccc cacctgccag cccatggggc tnggggggtag 84660
tgcaggcccc agagagacca agtgcaaagg agtacagctc attgcctctc cttcctcctg 84720
cagtacaaca agatcctgtt ggatatggaa accacctaca gcgtggccac tgtgtgccac 84780
ccgaatggca gctgcctgca gctcgagcca ggtgagagct catgtgcagg ctgagtgaga 84840
ggcgagggct gggactggca tggggcccgg gggtgctggg tgagagcaca gagttgggtt 84900
cccctcgctc ttggggtcag cgtgcccagg aaatgccctt tcttgttttc cacgaggggg 84960

gcttctctgc cccactgaga gccggcacct acttcatacc atgccccgat cagctgcccc 85020
tccctcagaa ccgccctctg cttaagggtg tccactctct cctgtcctct ctgcatgccg 85080
cccctcagag cagcgggatc tcaaagttat atttcatggg cttggactcc aaatgggggg 85140
aactcgggga cactagctcc ccccggcctc ctttcgtgac cctgcccttg acttcctcac 85200
cttctctgtc tttcctgagc ccctctccca gcatgtgact gataaggaaa ttgagtcaca 85260
cagcccctga aagcgccaga ctagaacctg agcctctgat tcctctcact tccctcacct 85320
accctgccac ttcctactgg atagaagtag acagctcttg actgtcctct tttctcccca 85380
ctggctggtc cttcttaccc cggcccgttt gaaagagctc acccccgaca caaggacccg 85440
cacacagata cctcccagct ccctctcaac ccacccttтc cagggttgga gaacttgagg 85500
cataaacttg cttccatgag gaatctccac ccagaaatgg gtctttctgg cccccagccc 85560
agctcccaca ttagaacaat gacaaatага aggggaaatg gaaaataaac aggagaaacg 85620
gttttcccag gacagggttt ggcctacaag ttgtggatgt gggtacccat gccaagtgtg 85680
aggggaggct ggccgggtgt ggtggctcat gcctctaatc ccagcacttt gggaggccaa 85740
ggtgagtaga tcacttgagg ccgggagttt gagaccagcc tggccaacat ggtgaaaccc 85800
catctgtact aaaaatacaa aagttagctg ggcgtggtgg tagatgcctg tagtcccagc 85860
tacttgggag gctgaggcat gagaatcgct tgagcccagc ctgggcaata cagcaagacc 85920
ccgtctctac aaataaaata caaaaaatta gttggatgtg gtggtgcatg cctgtagtcc 85980
tagctgctag ggaggctgag atggaaggat tgcttgagcc tgggaggtca aggctgcagt 86040
gagccgagat ggcgccactg cactccagcc tgggcaacag agtgagaccc tgtctcagaa 86100
aaaaaaaaaa aaaaaaaaag gagaggagag agactcaagc acgcccctca caggactgct 86160
gaggccctgc aggtgtctgc agcatgtggc cccaggccgg ggactctgta agccactgct 86220
ggagagccac tcccatcctt tctcccattt ctctagacct gctgcctata cagtcacttt 86280
tatgtggttt cgccaatttt attccagctc tgaaattctc tgagctcccc ttacaagcag 86340
aggtgagcta agggctggag ctcaaggcat tcaaacccct accagatctg acgaatgtga 86400
tggccacgtc ccggaaatat gaagacctgt tatgggcatg ggagggctgg cgagacaagg 86460
cggggagagc catcctccag ttttacccga aatacgtgga actcatcaac caggctgccc 86520
ggctcaatgg tgagtccctg ctgccaacat cactggcact tgggtccctt cattttcctc 86580
aaagaggtgc tgtgaaaccc caagcctagg aaaaggtaga tccctggagg aggcaggtaa 86640
tgtggtgttg ggagagcctg gctgtgtccc ctctgtaggc tatgtagatg caggggactc 86700
gtggaggtct atgtacgaga caccatccct ggagcaagac ctggagcggc tcttccagga 86760
gctgcagcca ctctacctca acctgcatgc ctacgtgcgc cgggccctgc accgtcacta 86820
cggggcccag cacatcaacc tggaggggcc cattcctgct cacctgctgg gtaagggcac 86880
atgtcgggcc ttgaggaggg taaagacgga ccacagtgtg agtgagggtt gggacagggc 86940
tgactagagg gtagggagca ggctggggac tgagagactc cagccctgtg ggggatggtt 87000
gcccaggctg gaggggggtg ggcgctggga gtggggagcc ccccacttgc atctggtgcc 87060
acattcactg cagatctatg tcgggcaagt caccatggat gggggaagaa gttaataatc 87120
ttgtccagga gaccacggca cccatcacaa cattgtgtga tcttagaggg cgaggaagag 87180
gctgtgagtg ggagctgggg aggctttgcc aagaggtggc ctgtgagcag ggcctcggaa 87240
gatgacaggg tttgacagat gggaagtggg ggatgagagg acagacgcag tgttcaggcc 87300
aagggaactg gaacaaagaa gaacctgaga atgtaaatct acttcaaccc tggaccctcc 87360
tttgccaagg gctgcaatct cagatgccct gaatgtgtga agtaggcggt gaggacagta 87420
agggatggta gggagtaagg caaagcagag gctactggtt ctctgtccct gatgggctgt 87480
taggaacact ttcctggagc agagagacca gacaggccct cagaccattt agaaactata 87540
agggaggccc cagaggacgg cctggctgtg ggtctagctc ccacacaggc tgggagtcca 87600
gccctcttca gcccctctct ggtgagacca aagaacatct ggtgatgtca cagtggacgt 87660
cagttcacca actgggagac acaggncccc gggaagaaaa gcaacatgcc cagcgtggcc 87720
tgggagctgg ggcagagctg gccttagaac tcagcccctg acaattggta aaaggggaaa 87780
ggggagcaac ctaacactga tgcgctctct gtctctctct ctggctctct ccctggctct 87840
ctccctcttc tctctcatgt tctctccatc actcatcgct ctaaccctct ctcactggtt 87900
tgcactttag acttcatctg atgtcagccg aagcttcacc tcacttggct aggaaagagc 87960
cttgagtcca aatctgtttc tgagccttcc attcatcctg agtttcttcc ttttcctctg 88020
tcgtggagaa ctaggctctt ttcttacact aaactcagag gcatcagcct ctncctgaag 88080
gagacggctg gttcctgtca gagttgctga gctgcagaca ccgacctcag gtggtgcgga 88140
ggggacatgg cagagtggct ggtgaagaga gcagcctgcc agcctttcaa tcccagccct 88200
gccacttagg agccgtgggc ccccggcgag gggggcgnng tncacttaac tctccagcct 88260
gtttccttta ctagccaatg ggaatcgtga cagtacctgg gtgcagacag gattgaaagt 88320
gaattcacac aatgttcttg gtgcagagcc aataagaggt ggccaccggg gtgtaggtgt 88380
tctggggacc tgtaatgtcc tcacatgtca gcagttgcta gtcacattgg tctccactgc 88440
tcacggacag tgaagaccac ctggattccc tgataaccag caaggccccc acctagagcc 88500
aggcagtaat gacctactgg gctggatatg cacaccaaag atgatgtgtg cctcaaagct 88560
tgcaaacagt aggtgctcaa gaaatgccac tatgattagc aggactggga tctggagcgc 88620

tcttcctgca ggagggcatt gagcctaagt aacatttgtc tttcctctct ctgccgtccc 88680
ccacactcgc ctccagggaa catgtgggcg cagacctggt ccaacatcta tgacttggtg 88740
gtgcccttcc cttcagcccc ctcgatggac accacagagg ctatgctaaa gcaggtccgc 88800
accagcccag gggcagggag gncccgccgg gantgggagg gaccctctga ttcaggagtt 88860
ccctccagtt tagccctccc ccgggatccc cacggcagca cgcagtctgn tccccggaac 88920
ccccagtttg ggcagaactc cctctngctt gcagggctgg acgcnccagg aggatgttta 88980
aggaggctga tgatttcttc acctnccctg gggctgctgc ccgtgcctcc tgagttctgg 89040
aacaagtcga tgctggagaa gccaaccgac gggcgggagg tggtctgcca cgcctcggcc 89100
tgggacttct acaacggcaa ggacttccgg tacatccagc tagggctcag gtctcgttcc 89160
tgagccccac gggcnaaggg aaatgaacca agcaaagggt ccactactgt cccccagctg 89220
gagccagcag ggcaggatgg ggacagggcc agagtttggg actgagtgtc tagagaggtg 89280
ttggcttctg gcaggaaaac cccatccgcc tgatggggac ttctgaagca cgcaacagct 89340
ctgtcagcct ggccgctggg aagtgctcaa ggtcccagtc ctgggtttga gcatggtagg 89400
ctgccccgcg tccctccttg ggagcagccc ctgcatggag ctggcctctc cctgggggca 89460
catgctgtga cacagggagg cacacgagga tgttgggtgc tctgtacaga tccactctca 89520
cccctgacag gctcagaagc tgccttcctt ggaggatggc gttttagtta cctattgctg 89580
tgcaaccaag caccccagag cttagcctta cgaaacaanc cagttgannt tttgcttatg 89640
attttatgtg ccaggaattc aggcagtaca cagtggaaat ggcctttctc tacagggccc 89700
cctctgttgg gggcagctct cacagccggg atggctcaat gggggccata cgtccagagc 89760
cccagttctg gctgtcggtt gaggtcctcg gtttttccca tgtggcagat gctggggcac 89820
atgttcccag tggcctcttg gctcacatgt tgggtgcttg gggtgagatg gctggaacgg 89880
ctggaggtgg gtcaggcatc tctccaggct agctcgggcg ccctcccagc gtggaatctg 89940
aggtgggcag atttacctgg cagccagcat cccccacagc aactactgac gcagccagtt 90000
ctcaaggcta ggtccaaaac tggcccagag tcacttctgc catgttttat tggctagaac 90060
aagtcacaag ttcacccaga ttcaagagaa gagaaaaaag tccctcccac ttgggagaag 90120
tggcaaagac catctgtcac agctgaagaa gtgtctctta caaggagaac agacacgggg 90180
agcctgaaac aaaacccgat gggattccct gggctgtgca ggcccttcca ggcatgagga 90240
ctcagccaca gggctngaga nggagacagg atctgggggа tgagagccct tgtggggtct 90300
tccctttтat ggggagtcag aggagaagct ggatagatcc ccagccttgt ggccaggatg 90360
ctgggcagct cttccttccc cctccccgat gagaatgaca gaaaaacagg attcacctga 90420
gccaaaaggc ttccagttag atccaagaga gaantttccc gcagtttgaa ttggtttgct 90480
aaacaacaag gaagggctgg gtgcggtggc tcacacctgt aatctcagca ctttgggaag 90540
ccgaggcagg aggtctgctt gagctcaggg gttcgagacc atcctgggca acatagcgag 90600
accccatttc ataaaaaata aataagtaaa tgagaacaag gaaggactga cgagagacgg 90660
tagaaccttc tggtttgggc nagctctgca gctgccattc atcctggcca taagaattct 90720
tggggtgaat aagtttgtcg ctgttgggcc gcatgagatg cagaatcgcc cactctcacc 90780
cctgacagaa acagttgttt ccttcaggga gcctccatct tgggagataa agcatgtgta 90840
catgggaacc cactggccac acattctcta gaaagtacac aatgtcccag tgcctctaga 90900
gcaagcactt tgtacagtca gaaagcaaca ggtggtgggg gctggagtca ttcaggaaaa 90960
tgggaggcag aggaatggcc tgaacggccc gatgctaggg gcttctgccc ccagattccc 91020
tcttacgcac actcagtggt tgcccttccc ctccctcccc acagtgctgt gtcccctgca 91080
tgctgcagtg ctggggtctg ccctgggtat agcaaggccc actgttccct tatgcccagg 91140
gcttctcact gtcctctccc aacaccctct cccccactcc actattccta ggatcaagca 91200
gtgcaccacc gtgaacttgg aggacctggt agtggcccac cacgaaatgg gccacatcca 91260
gtatttcatg cagtacaaag acttacctgt ggccttgagg gagggtgcca accccggctt 91320
ccatgaggcc attggggacg tgctagccct ctcagtgtct acgcccaagc acctgcacag 91380
tctcaacctg ctgagcagtg agggtggcag cgacggtgag agagaagcgg gaggccctgg 91440
tgggctgagg accaagaaag ggtggtgagc ttgggaggtg ggaaaggggc acttagtggc 91500
ccatgggcag aggtgtgggg cagagcaatc ggaaggaagg gagccaccca gaccatccca 91560
ggaggcaggt cacagggccc aaaaggtaca gcacccccac ccctccacca tcacaggcac 91620
accagggcca agccgctagg accctgggtc tgacagctgg gctcccttcc cttgcagagc 91680
atgacatcaa ctttctgatg aagatggccc ttgacaagat cgcctttatc cccttcagct 91740
acctcgtcga tcagtggcgc tggagggtat ttgatggaag catcaccaag gagaactata 91800
accaggagtg gtggagcctc aggttctgga acactcccac gggatgcggg ctgggggatc 91860
tctgcgagtg tctgcatgtg cctgggtgtc tggatgggcc agggtagggg agtgtgtgtg 91920
tgtgtgtaca ctattgtgtc tgtgcatatg atgtgtgtgg tgtaagtgat ggggaaaacg 91980
gggtgattgt gcacagaggc ccagcacgca ggagaatggg gtgcccagta tagccccaag 92040
tgcagggacc ctccctcaag tcaaaaatgc cacccccagc ctggttctcc ccaaactcat 92100
cttccaacat atattcccac tcgacaggct gaagtaccag ggcctctgcc ccccagtgcc 92160
caggactcaa ggtgactttg acccaggggc caagttccac attccttcta gcgtgcctta 92220
catcaggtaa cgggaaaggc aggagggcac attgtgaggg gcagtaccca cagctttgtg 92280

tttcaactgc ggccactgcc cggtccacaa gctctgtcag tcagggcaga cccgggggag 92340
ccggccgcac ggtgcaggtg cctgggccca ctcacactgc caaggctgat gggtttttc 92400
ttgaacattc ttttgatgag agtctgtacc atccaaacag ttgaacacag aaactcaacc 92460
taataattgg ctaatggtta ccagaccttg gttaagtagt taacattaac cacgactcat 92520
ggctggatca tgagctctgc actgttttgt tttgctttta aaacaagact gtgattcttt 92580
tactattatt gaacattgtc tgcgatacaa tttgaattgt acctggaagc ccttctagac 92640
actaaaatgt aggattggag atcggttaag gtgggaggca gggttgctgg ggcaagttac 92700
agtcacaggc tggggtcaga cagaactggg ttcaaactct gtctccnatt actttgtttc 92760
cttgagaaaa ttcctcaatt tctgtgagct tccatttcct gacctgtgaa ccccatttca 92820
caggatgcac gatggctaac ttcttagcat tctgtctcat acacagcctc ctcagggagg 92880
ggtggccagg accccactat tcatcactct ctagtggaat gtagctgcac actaggtctg 92940
caggtcacat ggccacagat gagtgtgccc aatgcagccc ctctccttct gtgtgccccg 93000
ggagagcact tgctgagggc tagcaaggct gtttgtgatc cgggaggctc cctgggaggc 93060
tgggggctag agagacctca ggctggagtt ccaggtgccc ccgggctaca ggtagcccag 93120
gccaccccca gagggctgtg gctgcctctg gccctggcct cccgtggttc ctggaagccc 93180
agcaagggca gggcccatgc ccaccttgcc tcctggcacc tgggatgatg ccagcacatc 93240
atcaagtgct aataactgat tgtgggatgg atgaagtctg tccccagagt ccaggaagag 93300
ggcatccctg gagcacctca gataggncct gacctagacg gtgctccaga gatgacactt 93360
aggancaggg ctcccgcctg cctgctggag tggtccctgg ggttcccagc cggcgctggc 93420
tctcacccgg gagccagctg gtgtgatggc tagcttccca gcttaatgca gacaattctc 93480
caaacagggg tgggcaaagg agacttggct gctctagaaa aacattccgg attctggcca 93540
gcagccttca caaagcactt ttaggaaaga ccagggaact aggtggtaca tgcttgcacc 93600
cagcattcaa agtgagaggc ctgtgccact ggctcaggac atttaaaacc tcttcagact 93660
ttaagctggg gagaatcctc cagccttgac tggcagattt ctaccaggga attcgtgatg 93720
ctttggataa gatcatgtag gactggcttc cctgcccaga ccaccctagt agatccacna 93780
cggcccattt ggccacacct tgcctctact tgcatatacc ccagggatgg agacctcact 93840
gcctccaaag ccacctgcca gatctcntga aggctctgcc ctgaccccat ggagcaggcc 93900
ctcctgagtt gtggaggcag ctctgtgggt gggaggcatc tacacaggca cggctaggaa 93960
gaggctcaga caatgctaag agctggggtg ggggagctca ccctgatagc tgtgggcaga 94020
gttgggggc cttggctctg ctgtgcgcat gtgacttagc acacatcaca tgtgatgtgc 94080
agaagggcct ggggcccagn tggcacaagg ccctcaacca actccgcccc gggccacggc 94140
ctcgctctgc tccaggtact ttgtcagctt catcatccag ttccagttcc acgaggcact 94200
gtgccaggca gctggccaca cgggcccccc gcacaagtgt gacatctacc agtccaagga 94260
ggccgggcag cgcctggcgn tgagtgtcct ccagccctcc tttgtttcca tngcntctgg 94320
cctgcgcccc tgggccttga ggggtctgtc cactggagct tttgtgggaa cacttgccat 94380
tttgagccgg gaactcccac ctgcagcgtg ggccaggcct gattgccatc tccttaggca 94440
cctggagccc tggggccctg ggacaagttt cagctgggag tgggtatgga gagtggatgt 94500
caggtggggg caagaggggc catgtccttc tgactctgcc tccctgtctc atgcctcccc 94560
aggaccgcca tgaagctggg cttcagntag gccgtggccg ganagccatg cagctgatca 94620
cgnnggccag cccnaacatg agcgcctcgg ccatgttgag ctacttcaag ccgctgctgg 94680
anctggctcc gcacggagaa cgagctgcat ggggagaagc tgggctggcc gcagtacaac 94740
tggacgnccg aactccggta ccgccaccca ccccacctcc agccttgggt cttaaccccc 94800
tccccaggct gggcagccat gcggctgacn ctncggagcc tggccctgcc ccgcaccctt 94860
gccctgccct gccctgccct gcccatgctg tctccttgct tcccgctcag ctcgctcaga 94920
agggcccctc ccagacagcg gccgcgtcag cttncctggg cctggacctg gatgcgcagg 94980
aggcccgcgt gggcncagnt ggctgctgct cttcctgggc atcgccctgn ctggtagcca 95040
ccctgggcct cagccagcgg ctncttncag catccgccac cgcagcctcc accggcactc 95100
ccacgggccc cagttcggct ccgaggtgga gctgagacac tcctgaggtg acccggctgg 95160
gtcggccctg cccaagggcc tcccaccaga gactgggatg ggaacactgg tgggcagctg 95220
aggacacacc ccacacccca gcccaccctg ctcctcctgc cctgtccctg tccccctccc 95280
ctcccagtcc tccagaccac cagccgcccc agccccttct cccagcacac ggctgcctga 95340
cactgagccc cacctctcca agtctctctg tgaatacaat taaaggtcct gccctcccca 95400
tctgagtctg tgtccctcac agggaagcca gggacaggga caggctgctt tcctgcctcc 95460
tggcagtcaa gtgggtcccg ttactaggtt tgttcctcca tcctccttca ggagccgggg 95520
aggatcccca gagctctgcc ccagcacctn cctggcntgg cgcctgnntc ttccctccag 95580
cccaggcagc ccgccactgt cctgccaccg caggcagccc ctgtctggcc caagcactga 95640
cccacgcgga ctctgggaag cagacatcct gggctgctgg cctcacattt ccactggcag 95700
tggagccttt ccctgctcca caaatggcca ggtcccccca ggggaaggct tccggctgtt 95760
atcggctgcc tcaggggggcg agtaccttgg agggcctgct tcaanggagg gtgccccctg 95820
gagggcacac accagcctag tgcttacctt ggctcctgcc tgtaccagct ccatgactct 95880
gctcgggtga acagccttgg ctctcagaca gccattctaa cactgccagt gcagaggggc 95940

ctcagacgct ggagtgtagc agtggctgca cctgcacagg gattagctgc cagcagccac 96000
cctgctggcg tcccagcaca cacctcctca ctccctgcat tggagggagt gtcattttaa 96060
gggacatttt tatgactttt atgtgtatgt ttatgtagaa atttggaaaa tacagaaaac 96120
tgtaaagaaa ataaaagccc tttatatcaa cgtcaagaga taagccctgt tgacgttttg 96180
gtgtacaact tgccggactt cttctcagca catgtgtatt ttaaatggga tcacaccata 96240
tttacagtca tacatccttt tatcacttca tacaactaga tctgtttttc tgatatttaa 96300
atgccagact tcgaacttgg ccaatagaag ttggtccatt tgctggggcc aggggctccc 96360
cagccacccg ggcccctctg tcaaaccctc agccctgagt ctcttctggg ctttgctgat 96420
gtcttcaccc tagctaggtc catctcccaa tatctgtccc ccttagtcca cagcttttgc 96480
cccccaatcc aggtgccgtg cgtgtctctg tgtgtccgtg tctgtgtgtg cgttgtacac 96540
aggcttggct gttacaggcc cattctgtaa ggcaggatgt ggggctgagg tattttagga 96600
ttgaaagagg gtggaagttt atgattacat aggacaatgg aatttataaa catgtcctct 96660
aaatggttcc gagtcatcta ccaatgaaga cttcttgaat tatccccccc tttcccagcc 96720
tgttttgaaa gctctttgtt taccagacaa aggtcatcaa tcatatgacc ccctttgcct 96780
ttttttttt tttaagatgg agtctcgctc ttgttgccca ggctagagtg cagtggtgtg 96840
atctcggctc actgtaacct ccacctcctg gctttcaagc gattctcctg cctcagcctt 96900
ccgagtagct gggattacag gcgcccacca ccatgcctgg ctaaattttt gtatttttag 96960
tagagatggg gtttaatcat gttggccagg ctggtctcga attcctgacc tcaggtgatc 97020
cacccacctc ggcctcccaa agtgctggaa ttacaggcat gagccaccat gcctggcccc 97080
cgtttcctat ttttatgaac cacagcggtt catgctgcct gtcagagctt ctgggccgcg 97140
tgaggtcacc agctttcaac acgcaaagga ctgcactgca gctgggggaa gagaaactcc 97200
acactgcatt ggcctggcca gccttaccct ctgggctttt gaaatagtat ctttttctg 97260
tttgttttca aacagagtct cgctctgtcg cccaggctgg agtgctggag tgcagtggcc 97320
tgatctcggc tcactgcaac ctccacctcc caagttcaag cgattctcct gcctcagcct 97380
cccgagtagc tgggctacta cttncaggcg cacgccgccn atgcccagct aatttctttt 97440
gtattttagt agagatggag tttcaccatg ttgcccaggc tggtcttgaa ctcctgagct 97500
caggcaatcc gcccgcctca gcctctcaaa gtgctaggat tancaggcgt gagccaccgc 97560
gcccggccca atagtatcat tctttagatg cctgcctctg cctccttggg tgagtgggga 97620
gaggcagggg atacctggaa agtagcagag gaagaggagg cggtaacagc aggaagaggg 97680
ccagcccagt gttttctact ggtggccctg aaggctgagc ccatccccgt gccgtgcctg 97740
ccaatgccgc tcttgggaga ccagctctca cctacgctag ccacaggtgg tggctgccag 97800
acagtttctc tgatccccac agccctcccc accctctacc ttcctctgtc tgcctaaccc 97860
ccttcccacc caccctggct tttaacataa gtgaaaaagt ggctaacccc acctctgcac 97920
ttatcacctg tgtgaccttg ggcagtttgt ttttgcagtc tgcattttct tttcttttct 97980
ttttttttt ttttttttt tgagatggag tctcgctctg tcacccaggc tggagtgcag 98040
tggcgtgatc tcggctcacc gcaagcttgg cctcctgggt tcacgccatt ctcctgcctc 98100
agcctcccga gtagctggga ctacaggcgc cagccaccac gcccggttaa ttttttgtat 98160
ttttagtaga gacagggttt caccgtgtta gccaggatgg tctcaatctc ctgacctcgt 98220
gatttgcctg cctcggcctc ccagagtgct gggattacag gcgtgagcca ccgcgcccag 98280
cctgcgtttt ctttcttacg gttcttatca accctctcag cgttgccatg aagatgaaat 98340
gagatgatgt acaaagtcct agtagagtgt cttctcttta taatgaatgc atcgtctcct 98400
gagaaagcta gtttcataac aaccccagat cagccaagtc cagatcagcc ctctcacttg 98460
agacaggaag aggacccggg gcaactgggt gccggagctg gactgaaaac tcccatctcc 98520
cagctgcctt ccaggaactt ccccaccaca cgtccttgca caaccagtca actgtcttct 98580
tactgggagc acacagagct cgtccactgg gggcccacag cttgcctcag ttccaggagt 98640
actcagccat ctccctgtgt cgcctctccc tcatcatccc tccccatgtc acatctccct 98700
cagcctctcc gtgttacctc tccctcatcc tccctcccca tgttgcctct ccctcatcct 98760
ctctctcatc ctccctgccc gtgtcgcctc ttcctcatcc tctccctcat cctctccctc 98820
ttcctctct 98829

<210> 4
<211> 1306
<212> PRT
<213> Homo Sapiens

<400> 4
Met Gly Ala Ala Ser Gly Arg Arg Gly Pro Gly Leu Leu Leu Pro Leu
 1 5 10 15
Pro Leu Leu Leu Leu Leu Pro Pro Gln Pro Ala Leu Ala Leu Asp Pro
 20 25 30
Gly Leu Gln Pro Gly Asn Phe Ser Ala Asp Glu Ala Gly Ala Gln Leu

```
        35                  40                  45
Phe Ala Gln Ser Tyr Asn Ser Ser Ala Glu Gln Val Leu Phe Gln Ser
    50                  55                  60
Val Ala Ala Ser Trp Ala His Asp Thr Asn Ile Thr Ala Glu Asn Ala
65                  70                  75                  80
Arg Arg Gln Glu Glu Ala Ala Leu Leu Ser Gln Glu Phe Ala Glu Ala
                85                  90                  95
Trp Gly Gln Lys Ala Lys Glu Leu Tyr Glu Pro Ile Trp Gln Asn Phe
            100                 105                 110
Thr Asp Pro Gln Leu Arg Arg Ile Ile Gly Ala Val Arg Thr Leu Gly
        115                 120                 125
Ser Ala Asn Leu Pro Leu Ala Lys Arg Gln Gln Tyr Asn Ala Leu Leu
    130                 135                 140
Ser Asn Met Ser Arg Ile Tyr Ser Thr Ala Lys Val Cys Leu Pro Asn
145                 150                 155                 160
Lys Thr Ala Thr Cys Trp Ser Leu Asp Pro Asp Leu Thr Asn Ile Leu
                165                 170                 175
Ala Ser Ser Arg Ser Tyr Ala Met Leu Leu Phe Ala Trp Glu Gly Trp
            180                 185                 190
His Asn Ala Ala Gly Ile Pro Leu Lys Pro Leu Tyr Glu Asp Phe Thr
        195                 200                 205
Ala Leu Ser Asn Glu Ala Tyr Lys Gln Asp Gly Phe Thr Asp Thr Gly
    210                 215                 220
Ala Tyr Trp Arg Ser Trp Tyr Asn Ser Pro Thr Phe Glu Asp Asp Leu
225                 230                 235                 240
Glu His Leu Tyr Gln Gln Leu Glu Pro Leu Tyr Leu Asn Leu His Ala
                245                 250                 255
Phe Val Arg Arg Ala Leu His Arg Arg Tyr Gly Asp Arg Tyr Ile Asn
            260                 265                 270
Leu Arg Gly Pro Ile Pro Ala His Leu Leu Gly Asp Met Trp Ala Gln
        275                 280                 285
Ser Trp Glu Asn Ile Tyr Asp Met Val Val Pro Phe Pro Asp Lys Pro
    290                 295                 300
Asn Leu Asp Val Thr Ser Thr Met Leu Gln Gln Gly Trp Asn Ala Thr
305                 310                 315                 320
His Met Phe Arg Val Ala Glu Glu Phe Phe Thr Ser Leu Glu Leu Ser
                325                 330                 335
Pro Met Pro Pro Glu Phe Trp Glu Gly Ser Met Leu Glu Lys Pro Ala
            340                 345                 350
Asp Gly Arg Glu Val Val Cys His Ala Ser Ala Trp Asp Phe Tyr Asn
        355                 360                 365
Arg Lys Asp Phe Arg Ile Lys Gln Cys Thr Arg Val Thr Met Asp Gln
    370                 375                 380
Leu Ser Thr Val His His Glu Met Gly His Ile Gln Tyr Tyr Leu Gln
385                 390                 395                 400
Tyr Lys Asp Leu Pro Val Ser Leu Arg Arg Gly Ala Asn Pro Gly Phe
                405                 410                 415
His Glu Ala Ile Gly Asp Val Leu Ala Leu Ser Val Ser Thr Pro Glu
            420                 425                 430
His Leu His Lys Ile Gly Leu Leu Asp Arg Val Thr Asn Asp Thr Glu
        435                 440                 445
Ser Asp Ile Asn Tyr Leu Leu Lys Met Ala Leu Glu Lys Ile Ala Phe
    450                 455                 460
Leu Pro Phe Gly Tyr Leu Val Asp Gln Trp Arg Trp Gly Val Phe Ser
465                 470                 475                 480
Gly Arg Thr Pro Pro Ser Arg Tyr Asn Phe Asp Trp Trp Tyr Leu Arg
                485                 490                 495
Thr Lys Tyr Gln Gly Ile Cys Pro Pro Val Thr Arg Asn Glu Thr His
            500                 505                 510
Phe Asp Ala Gly Ala Lys Phe His Val Pro Asn Val Thr Pro Tyr Ile
        515                 520                 525
```

```
Arg Tyr Phe Val Ser Phe Val Leu Gln Phe Gln Phe His Glu Ala Leu
    530                 535                 540
Cys Lys Glu Ala Gly Tyr Glu Gly Pro Leu His Gln Cys Asp Ile Tyr
545                 550                 555                 560
Arg Ser Thr Lys Ala Gly Ala Lys Leu Arg Lys Val Leu Gln Ala Gly
                565                 570                 575
Ser Ser Arg Pro Trp Gln Glu Val Leu Lys Asp Met Val Gly Leu Asp
            580                 585                 590
Ala Leu Asp Ala Gln Pro Leu Leu Lys Tyr Phe Gln Pro Val Thr Gln
        595                 600                 605
Trp Leu Gln Glu Gln Asn Gln Gln Asn Gly Glu Val Leu Gly Trp Pro
    610                 615                 620
Glu Tyr Gln Trp His Pro Pro Leu Pro Asp Asn Tyr Pro Glu Gly Ile
625                 630                 635                 640
Asp Leu Val Thr Asp Glu Ala Glu Ala Ser Lys Phe Val Glu Glu Tyr
                645                 650                 655
Asp Arg Thr Ser Gln Val Val Trp Asn Glu Tyr Ala Glu Ala Asn Trp
            660                 665                 670
Asn Tyr Asn Thr Asn Ile Thr Thr Glu Thr Ser Lys Ile Leu Leu Gln
        675                 680                 685
Lys Asn Met Gln Ile Ala Asn His Thr Leu Lys Tyr Gly Thr Gln Ala
    690                 695                 700
Arg Lys Phe Asp Val Asn Gln Leu Gln Asn Thr Thr Ile Lys Arg Ile
705                 710                 715                 720
Ile Lys Lys Val Gln Asp Leu Glu Arg Ala Ala Leu Pro Ala Gln Glu
                725                 730                 735
Leu Glu Glu Tyr Asn Lys Ile Leu Leu Asp Met Glu Thr Thr Tyr Ser
            740                 745                 750
Val Ala Thr Val Cys His Pro Asn Gly Ser Cys Leu Gln Leu Glu Pro
        755                 760                 765
Asp Leu Thr Asn Val Met Ala Thr Ser Arg Lys Tyr Glu Asp Leu Leu
    770                 775                 780
Trp Ala Trp Glu Gly Trp Arg Asp Lys Ala Gly Arg Ala Ile Leu Gln
785                 790                 795                 800
Phe Tyr Pro Lys Tyr Val Glu Leu Ile Asn Gln Ala Ala Arg Leu Asn
                805                 810                 815
Gly Tyr Val Asp Ala Gly Asp Ser Trp Arg Ser Met Tyr Glu Thr Pro
            820                 825                 830
Ser Leu Glu Gln Asp Leu Glu Arg Leu Phe Gln Glu Leu Gln Pro Leu
        835                 840                 845
Tyr Leu Asn Leu His Ala Tyr Val Arg Arg Ala Leu His Arg His Tyr
    850                 855                 860
Gly Ala Gln His Ile Asn Leu Glu Gly Pro Ile Pro Ala His Leu Leu
865                 870                 875                 880
Gly Asn Met Trp Ala Gln Thr Trp Ser Asn Ile Tyr Asp Leu Val Val
                885                 890                 895
Pro Phe Pro Ser Ala Pro Ser Met Asp Thr Thr Glu Ala Met Leu Lys
            900                 905                 910
Gln Gly Trp Thr Pro Arg Arg Met Phe Lys Glu Ala Asp Asp Phe Phe
        915                 920                 925
Thr Ser Leu Gly Leu Leu Pro Val Pro Pro Glu Phe Trp Asn Lys Ser
    930                 935                 940
Met Leu Glu Lys Pro Thr Asp Gly Arg Glu Val Val Cys His Ala Ser
945                 950                 955                 960
Ala Trp Asp Phe Tyr Asn Gly Lys Asp Phe Arg Ile Lys Gln Cys Thr
                965                 970                 975
Thr Val Asn Leu Glu Asp Leu Val Val Ala His His Glu Met Gly His
            980                 985                 990
Ile Gln Tyr Phe Met Gln Tyr Lys Asp Leu Pro Val Ala Leu Arg Glu
        995                 1000                1005
Gly Ala Asn Pro Gly Phe His Glu Ala Ile Gly Asp Val Leu Ala Leu
```

```
    1010                1015                1020
Ser Val Ser Thr Pro Lys His Leu His Ser Leu Asn Leu Leu Ser Ser
1025                1030                1035                1040
Glu Gly Gly Ser Asp Glu His Asp Ile Asn Phe Leu Met Lys Met Ala
                1045                1050                1055
Leu Asp Lys Ile Ala Phe Ile Pro Phe Ser Tyr Leu Val Asp Gln Trp
            1060                1065                1070
Arg Trp Arg Val Phe Asp Gly Ser Ile Thr Lys Glu Asn Tyr Asn Gln
        1075                1080                1085
Glu Trp Trp Ser Leu Arg Leu Lys Tyr Gln Gly Leu Cys Pro Pro Val
    1090                1095                1100
Pro Arg Thr Gln Gly Asp Phe Asp Pro Gly Ala Lys Phe His Ile Pro
1105                1110                1115                1120
Ser Ser Val Pro Tyr Ile Arg Tyr Phe Val Ser Phe Ile Ile Gln Phe
                1125                1130                1135
Gln Phe His Glu Ala Leu Cys Gln Ala Ala Gly His Thr Gly Pro Leu
            1140                1145                1150
His Lys Cys Asp Ile Tyr Gln Ser Lys Glu Ala Gly Gln Arg Leu Ala
        1155                1160                1165
Thr Ala Met Lys Leu Gly Phe Ser Arg Pro Trp Pro Glu Ala Met Gln
    1170                1175                1180
Leu Ile Thr Gly Gln Pro Asn Met Ser Ala Ser Ala Met Leu Ser Tyr
1185                1190                1195                1200
Phe Lys Pro Leu Leu Asp Trp Leu Arg Thr Glu Asn Glu Leu His Gly
                1205                1210                1215
Glu Lys Leu Gly Trp Pro Gln Tyr Asn Trp Thr Pro Asn Ser Ala Arg
            1220                1225                1230
Ser Glu Gly Pro Leu Pro Asp Ser Gly Arg Val Ser Phe Leu Gly Leu
        1235                1240                1245
Asp Leu Asp Ala Gln Gln Ala Arg Val Gly Gln Trp Leu Leu Leu Phe
    1250                1255                1260
Leu Gly Ile Ala Leu Leu Val Ala Thr Leu Gly Leu Ser Gln Arg Leu
1265                1270                1275                1280
Phe Ser Ile Arg His Arg Ser Leu His Arg His Ser His Gly Pro Gln
                1285                1290                1295
Phe Gly Ser Glu Val Glu Leu Arg His Ser
            1300                1305
```

<210> 5
<211> 8878
<212> DNA
<213> Homo Sapiens

<400> 5

```
gaattcatgc ccctttgaa atagacttat gtcattgtca gaaaacataa gcatttatgg 60
tatatcatta atgagtcacg attttagtgg ttgccttgtg agtaggtcaa atttactaag 120
cttagatttg ttttctcaca tattctttcg gagcttgtgt agtttccaca ttaatttacc 180
agaaacaaga tacacactct ctttgaggag tgccctaact tcccatcatt ttgtccaatt 240
aaatgaattg aagaaattta atgtttctaa actagaccaa caaagaataa tagttgtatg 300
acaagtaaat aagctttgct gggaagatgt tgcttaaatg ataaaatggt tcagccaaca 360
agtgaaccaa aaattaaata ttaactaagg aaaggtaacc atttctgaag tcattcctag 420
cagaggactc agatatatat aggattgaag atctctcagt taagtctaca tgaaaaggat 480
ggtttcttgg agcttccaca aacttaaaac catgaaacat ctattattgc tactattgtg 540
tgtttttcta gttaagtccc aaggtgtcaa cgacaatgag gaggtgaatt ttttaaagca 600
ttattatatt attagtagta ttattaatat aagatgtaac ataatcatat tatgtgctta 660
ttttaatgaa attagcattg cttatagtta tgaaatggaa ttgttaacct ctgacttatt 720
gtatttaaag aatgtttcat agtatttctt atataaaaac aaagtaattt cttgttttct 780
agtttatcac ctttgttttc ttaagatgag gatggcttag ctaatgtaag atgtgttttt 840
ctcacttgct attctgagta ctgtgatttt catttacttc tagcaataca ggattacaat 900
taagaggaca agatctgaaa atctcacaaa ctataaaata ataaaagagc agaatttaa 960
```

gataaaagaa actggtggta ggtagattgt tctttggtga aggaaggtaa tatatattgt 1020
tactgagatt actatttata aaaattataa ctaagcctaa aagcaaaata catcaagtgt 1080
aatgatagaa aatgaaatat tgctttttc agatgaaaag ttcaaattag agttagtgtg 1140
tattgttatt attaatagtt atgaaacacg gttcagtcta atttatttat ttgtagaaca 1200
gtttgtcctc aactattatt tttgctgact tattgctgtt aatttgcagt tactaaaaat 1260
acagaaatgc atttaggaca atggatattt aagaaattta aattttatca tcaaacgtat 1320
catggccaaa tttcttacat atagcatagt atcattaaac tagaaataag aatacacaat 1380
aatatttaaa tgaagtgatt catttcggat cattattgag tttcaaggga acttgagtgt 1440
tgtacttatc agactctaca tgtaagaaca tatagttaat ctggttgtgt gtgtaaaaac 1500
atatggttaa tctggttaag tctggttaat catattaggt aagaaaaatg taaagaatgt 1560
gtaagacgaa atttttgtaa agtactctgc aaagcacttt cacatttctg cttatcaact 1620
aaacctcaca gagatagttt aatagtttag gctttaaaat ggattttgat tattcaacaa 1680
gtggccttca taatttcttt aagtgttttt ctttaagtat atactttctt taaatatttt 1740
ttaaaatttc cttttctcta gtaaagccag accatccatg ctacctctct agtggcactc 1800
tgaaataaaa agaaaatagt tttctctgtt ataattgtat ttgtaataag cagatgaatc 1860
acatttctta aaatttgttt tagagagggt aagctctgac taggaccatg acttcaatgt 1920
gaaatatgta tatatcctcc gaatctttac atattaagaa tgtatatagt caactggtta 1980
aacaggaaaa tctggaacag cctggctggg ttttaatctt agcaccatcc tactaaatgt 2040
taaataatat tataatctaa tgaataaatg acaatgcaat tccaaataga gttcatctga 2100
tgacttctag actcacaaaa ttgcaagaga gctcagttgt tgctcagttg ttccaaatca 2160
tgtcgtttgt taatttgtaa ttaagctcca aaggatgtat agctactgac aaaaaaaaaa 2220
atgagaatgt agttaatcca aatcaaaact ttcctattgc aatgcgtatt ttctgcttca 2280
ttatccttta atataatatt ttaagttagc aagtaatttt aattacaatg cacaagcctt 2340
gagaattatt ttaaatataa gaaaatcata atgtttgata aagaaatcat gtaagaaatt 2400
tcaagataat ggtttaacaa ataattttgt tgatagaaga taagactaaa agtgaaattc 2460
gaagtggaga ggacacttaa actgtagtac ttgttatgtg tgattccagt aaaaatagta 2520
atgagcactt attattgcca agtactgttc tgagggtacc atatgcaata agttatttaa 2580
tccttacaat aatcttgtaa ggcagattca aactatcatt acacttattt tacagatgag 2640
aaaactgggg cacagataaa gcaacttgcc caaggtctca tagctgtaag tcaaccctac 2700
ggtcaagacc tacaagtagc cgagctccag agtacattat gagggtcaaa gattgtctta 2760
ttacaaataa attccaagta gaatcaacct ttaataagtc tttaatgtct cttaaatatg 2820
tttatatagg agtctaatca ccaattcaca aaaatgaaag tagggaaatg attaacaata 2880
atcataggaa tctaacaatc caagtggctt gagaatattc attcttcttg acagtataga 2940
ttctttacaa tttcgtaagt tccaatgtat gttttaggaa tatgaggtca ttactattca 3000
taatctgata cagctttatc ctaaggcctc tctttaaaaa ctacactgca tcatagcttt 3060
tttgtgcagt tggtctttct actgttactg aacagtaagc aacctacaga ttcactatca 3120
ccaaccagcc agttgatgga tcttaagcaa attatcaagc ttgtgataac ctaaattata 3180
aaatgagggt gttggaatag ttacattcca aatcttctat aacactctgt attatatttc 3240
tgcctcattc cttgtagggt ttcttcagtg cccgtggtca tcgacccctt gacaagaaga 3300
gagaagaggc tcccagcctg aggcctgccc caccgcccat cagtggaggt ggctatcggg 3360
ctcgtccagc caaagcagct gccactcaaa agaaagtaga aagaaaagcc cctgatgctg 3420
gaggctgtct tcacgctgac ccagacctgg tgggtgcact gatgtttctt gcagtggtgg 3480
ctctctcatg cagagaaagc ctgtagtcat ggcagtctgc taatgtttca ctgacccaca 3540
ttaccatcac tgttattttg tttgtttatt ttggaaataa aattcaaaac ataaacatat 3600
tgggcctttg gtttaggctt tctttcttgt tttctttggt ctgggcccaa aatttcaaat 3660
taggatatgt gggtgccacc tttccatttg tattttgcca ctgcctttgt ttagttggta 3720
aaatttttcat agcccaatta tatttttttct ggggtaagta atattttaaa tctctatgag 3780
agtatgatga tgactttcga atttctggtc ttacagaaaa ccaaataata aatttttatg 3840
ttggctaatc gtatcgctga attttcctat gtgctatttt aacaaatgtc catgacccaa 3900
atccttcatc taatgcctgc tattttcttt gtttttaggg ggtgttgtgt cctacaggat 3960
gtcagttgca agaggctttg ctacaacagg aaaggccaat cagaaatagt gttgatgagt 4020
taaataacaa tgtggaagct gtttcccaga cctcctcttc ttcctttcag tacatgtatt 4080
tgctgaaaga cctgtggcaa aagaggcaga agcaagtaaa aggtagatat ccttgtgctt 4140
tccattcgat tttcagctat aaaattggaa ccgttagact gccacgagaa tgcatggttg 4200
tgagaagatt aacatttctg ggttagtgaa tagcattcat acgcttttgg gcaccttccc 4260
ctgcaacttg ccagataagc actattcagc tcttattccc agtctgacat cagcaagtgt 4320
gattttctat gaaaaattct actatgactc cttattttaa gtatacaaga aacttgtgac 4380
tcagaagata atatttacag agtggaaaaa aacccctagc atttatagtt ttaacatttg 4440
aggttttgaa tgagagagtt atccataata tattcaattg tgttgtggat aatgacacct 4500
aacctgtgaa tcttgaggtc agaatgttga gtgctgttga cttggtggtc aggaaacagc 4560
tagtgcgtga gcctggcaca ggcatctcag tgagtagcat acccacagtt ggaaattttt 4620

caaagaaatc aaaggaatca tgacatctta taaatttcaa ggttctgcta tacttatgtg 4680
aaatggataa ataaatcaag catatccact ctgtaagatt gaacttctca gatggaagac 4740
cccaatactg ctttctcctc ttttccctca ccaaagaaat aaacaaccta tttcatttat 4800
tactggacac aatctttagc gtatacctat ggtaaattac tagtatggtg gttaggattt 4860
atgttaattt gtatatgtca tgcgccaaat catttccact aaatatgact atatatcata 4920
actgcttggt gatagctcag tgtttaatag tttattctca gaaaatcaaa attgtatagt 4980
taaatacatt agttttatga ggcaaaaatg ctaactattt ctacataatt tcatttttcc 5040
agataatgaa aatgtagtca atgagtactc ctcagaactg gaaaagcacc aattatatat 5100
agatgagact gtgaatagca atatcccaac taaccttcgt gtgcttcgtt caatcctgga 5160
aaacctgaga agcaaaatac aaaagttaga atctgatgtc tcagctcaaa tggaatattg 5220
tcgcacccca tgcactgtca gttgcaatat tcctgtggtg tctggcaaag gtaactgatt 5280
cataaacata tttttagaga gttccagaag aactcacaca ccaaaaataa gagaacaaca 5340
acaacaacaa aaatgctaag tggattttcc caacagatca taatgacatt acagtacatc 5400
ataaaaatat ccttagccag ttgtgttttg gactggcctg gtgcatttgc tggttttgat 5460
gagcaggatg gggcacaggt agtcccaggg gtggctgatg tgtgcatctg cgtactggct 5520
tgaacagatg gcagaaccac agatagatgt agaagtttct ccattttgtg tgttctggga 5580
gctcatggat attccaggac acaaaaggtg gagaagagct ttgttcatcc tcttagcaga 5640
taaacgtcct caaaactggg ttggacttac taaagtaaaa tgaaaatcta atatttgtta 5700
tattattttc aaaggtctat aataacacac tccttagtaa cttatgtaat gttattttaa 5760
agaattggtg actaaataca aagtaattat gtcataaacc cctgaacata atgttgtctt 5820
acatttgcag aatgtgagga aattatcagg aaaggaggtg aaacatctga aatgtatctc 5880
attcaacctg acagttctgt caaaccgtat agagtatact gtgacatgaa tacagaaaat 5940
ggaggtaagc tttcgacagt tgttgacctg ttgatctgta attatttgga taccgtaaaa 6000
tgccaggaaa caaggccagg tgtggtggct catacctgta attccagcac cttgggaggc 6060
caaagtgggc tgatagcttg agcctaggag tttgaaacta gcctgggcaa cataatgaga 6120
ccctaactct acaaaaaaaa aaaaaatacc aaaaaaaaaa aaaaaatcag ctgtgttggt 6180
agtatgtgcc tgtagtccca gctatccagg aggctgagat gggagatcac ctgagcccac 6240
aacctggagt cttgatcatg ctactgaact gtagcctggg caacagagga tagtgagatc 6300
ctgtctcaaa aaaaaaaatt aattaaaaag ccaggaaaca agacttagct ctaacatcta 6360
acatagctga caaaggagta atttgatgtg gaattcaacc tgatatttaa aagttataaa 6420
atatctataa ttcacaattt ggggtaagat aaagcacttg cagtttccaa agattttaca 6480
agtttacctc tcatatttat ttccttattg tgtctatttt agagcaccaa atatatacta 6540
aatggaatgg acaggggatt cagatattat tttcaaagtg acattatttg ctgttggtta 6600
atatatgctc tttttgtttc tgtcaaccaa aggatggaca gtgattcaga accgtcaaga 6660
cggtagtgtt gactttggca ggaaatggga tccatataaa cagggatttg gaaatgttgc 6720
aaccaacaca gatgggaaga attactgtgg cctaccaggt aacgaacagg catgcaaaat 6780
aaaatcattc tatttgaaat gggatttttt ttaattaaaa aacattcatt gttggaagcc 6840
tgttttaggc agttaagagg agtttcctga caaaaatgtg gaagctaaag ataagggaag 6900
aaaggcagtt tttagtttcc caaaatttta tttttggtga gagattttat tttgtttttc 6960
ttttaggtga atattggctt ggaaatgata aaattagcca gcttaccagg atgggaccca 7020
cagaactttt gatagaaatg gaggactgga aaggagacaa agtaaaggct cactatggag 7080
gattcactgt acagaatgaa gccaacaaat accagatctc agtgaacaaa tacagaggaa 7140
cagccggtaa tgccctcatg gatggagcat ctcagctgat gggagaaaac aggaccatga 7200
ccattcacaa cggcatgttc ttcagcacgt atgacagaga caatgacggc tggtatgtgt 7260
ggcactcttt gctcctgctt taaaaatcac actaatatca ttactcagaa tcattaacaa 7320
tatttttaat agctaccact tcctgggcac ttactgtcag ccactgtcct aagctcttta 7380
tgcatcactc gaaagcattt caactataag gtagacattc ttattctcat tttacagatg 7440
agatttagag agattacgtg atttgtccaa tgtcacacaa ctacccagag ataaaactag 7500
aatttgagca cagttacttt ctgaataatg agcatttaga taaataccta tatctctata 7560
ttctaaagtg tgtgtgaaaa ctttcatttt catttccagg gttctctgat actaagggtt 7620
gtaaaagcta ttattccagt ataaagtaac aaacacagtc cctagatgga ttgccacaaa 7680
ggcccagtta tctctctttc ttgctatagg gcacaggagg tctttggtgt attagtgtga 7740
ctctatgtat agcacccaaa ggaaagacta ctgtgcacac gagtgtagca gtcttttatg 7800
ggtaatctgc aaaacgtaac ttgaccaccg tagttctgtt tctaataacg ccaaacacat 7860
tttctttcag gttaacatca gatcccagaa aacagtgttc taaagaagac ggtggtggat 7920
ggtggtataa tagatgtcat gcagccaatc caaacggcag atactactgg ggtggacagt 7980
acacctggga catggcaaag catggcacag atgatggtgt agtatggatg aattggaagg 8040
ggtcatggta ctcaatgagg aagatgagta tgaagatcag gcccttcttc ccacagcaat 8100
agtccccaat acgtagattt ttgctcttct gtatgtgaca acatttttgt acattatgtt 8160
attggaattt tctttcatac attatattcc tctaaaactc tcaagcagac gtgagtgtga 8220
ctttttgaaa aaagtatagg ataaattaca ttaaaatagc acatgatttt cttttgtttt 8280

cttcatttct cttgctcacc caagaagtaa caaaagtata gttttgacag agttggtgtt 8340
cataatttca gttctagttg attgcgagaa ttttcaaata aggaagaggg gtcttttatc 8400
cttgtcgtag gaaaaccatg acggaaagga aaaactgatg tttaaaagtc cacttttaaa 8460
actatattta tttatgtagg atctgtcaaa gaaaacttcc aaaaagattt attaattaaa 8520
ccagactctg ttgcaataag ttaatgtttt cttgttttgt aatccacaca ttcaatgagt 8580
taggctttgc acttgtaagg aaggagaagc gttcacaacc tcaaatagct aataaaccgg 8640
tcttgaatat ttgaagattt aaaatctgac tctaggacgg gcacggtggc tcacgactat 8700
aatcccaaca ctttgggagg ctgaggcggg cggtcacaag gtcaggagtt caagaccagc 8760
ctgaccaata tggtgaaacc ccatctctac taaaaataca aaaattagcc aggcgtggtg 8820
gcaggtgcct gtaggtccca gctagcctgt gaggtggaga ttgcattgag ccaagatc 8878

<210> 6
<211> 491
<212> PRT
<213> Homo Sapiens

<400> 6

```
Met Lys Arg Met Val Ser Trp Ser Phe His Lys Leu Lys Thr Met Lys
 1               5                  10                  15
His Leu Leu Leu Leu Leu Leu Cys Val Phe Leu Val Lys Ser Gln Gly
            20                  25                  30
Val Asn Asp Asn Glu Glu Gly Phe Phe Ser Ala Arg Gly His Arg Pro
        35                  40                  45
Leu Asp Lys Lys Arg Glu Glu Ala Pro Ser Leu Arg Pro Ala Pro Pro
    50                  55                  60
Pro Ile Ser Gly Gly Gly Tyr Arg Ala Arg Pro Ala Lys Ala Ala Ala
65                  70                  75                  80
Thr Gln Lys Lys Val Glu Arg Lys Ala Pro Asp Ala Gly Gly Cys Leu
                85                  90                  95
His Ala Asp Pro Asp Leu Gly Val Leu Cys Pro Thr Gly Cys Gln Leu
            100                 105                 110
Gln Glu Ala Leu Leu Gln Gln Glu Arg Pro Ile Arg Asn Ser Val Asp
        115                 120                 125
Glu Leu Asn Asn Asn Val Glu Ala Val Ser Gln Thr Ser Ser Ser Ser
    130                 135                 140
Phe Gln Tyr Met Tyr Leu Leu Lys Asp Leu Trp Gln Lys Arg Gln Lys
145                 150                 155                 160
Gln Val Lys Asp Asn Glu Asn Val Val Asn Glu Tyr Ser Ser Glu Leu
                165                 170                 175
Glu Lys His Gln Leu Tyr Ile Asp Glu Thr Val Asn Ser Asn Ile Ala
            180                 185                 190
Thr Asn Leu Arg Val Leu Arg Ser Ile Leu Glu Asn Leu Arg Ser Lys
        195                 200                 205
Ile Gln Lys Leu Glu Ser Asp Val Ser Ala Gln Met Glu Tyr Cys Arg
    210                 215                 220
Thr Pro Cys Thr Val Ser Cys Asn Ile Pro Val Val Ser Gly Lys Glu
225                 230                 235                 240
Cys Glu Glu Ile Ile Arg Lys Gly Gly Glu Thr Ser Glu Met Tyr Leu
                245                 250                 255
Ile Gln Pro Asp Ser Ser Val Lys Pro Tyr Arg Val Tyr Cys Asp Met
            260                 265                 270
Asn Thr Glu Asn Gly Gly Trp Thr Val Ile Gln Asn Arg Gln Asp Gly
        275                 280                 285
Ser Val Asp Phe Gly Arg Lys Trp Asp Pro Tyr Lys Gln Gly Phe Gly
    290                 295                 300
Asn Val Ala Thr Asn Thr Asp Gly Lys Asn Tyr Cys Gly Leu Pro Gly
305                 310                 315                 320
Glu Tyr Trp Leu Gly Asn Asp Lys Ile Ser Gln Leu Thr Arg Met Gly
                325                 330                 335
Pro Thr Glu Leu Leu Ile Glu Met Glu Asp Trp Lys Gly Asp Lys Val
            340                 345                 350
```

```
Lys Ala His Tyr Gly Gly Phe Thr Val Gln Asn Glu Ala Asn Lys Tyr
        355                 360                 365
Gln Ile Ser Val Asn Lys Tyr Arg Gly Thr Ala Gly Asn Ala Leu Met
    370                 375                 380
Asp Gly Ala Ser Gln Leu Met Gly Glu Asn Arg Thr Met Thr Ile His
385                 390                 395                 400
Asn Gly Met Phe Phe Ser Thr Tyr Asp Arg Asp Asn Asp Gly Trp Leu
                405                 410                 415
Thr Ser Asp Pro Arg Lys Gln Cys Ser Lys Glu Asp Gly Gly Gly Trp
            420                 425                 430
Trp Tyr Asn Arg Cys His Ala Ala Asn Pro Asn Gly Arg Tyr Tyr Trp
        435                 440                 445
Gly Gly Gln Tyr Thr Trp Asp Met Ala Lys His Gly Thr Asp Asp Gly
    450                 455                 460
Val Val Trp Met Asn Trp Lys Gly Ser Trp Tyr Ser Met Arg Lys Met
465                 470                 475                 480
Ser Met Lys Ile Arg Pro Phe Phe Pro Gln Gln
                485                 490
```

<210> 7
<211> 17
<212> DNA
<213> Homo sapiens

<400> 7
aatggaacgc agagatg 17

<210> 8
<211> 31
<212> DNA
<213> Homo sapiens

<400> 8
tgcaaatggg tgtgacgcgg ttccagatgt g 31

<210> 9
<211> 31
<212> DNA
<213> Homo sapiens

<400> 9
gaatgtgatg gccacgtccc ggaaatatga a 31

<210> 10
<211> 31
<212> DNA
<213> Homo sapiens

<400> 10
tgagactgtg aatagtaata tcccaactaa c 31

<210> 11
<211> 31
<212> DNA
<213> Homo sapiens

<400> 11
catggtactc aatgaagaag atgagtatga a 31